Value Added Assessment of Teacher Preparation in Louisiana: 2004 - 2006

George H. Noell, Ph.D.
Bethany A. Porter, M.A.
R. Maria Patt, M.A.

Department of Psychology
Louisiana State University

October 24, 2007

Acknowledgements

This report is based upon data provided by the Division of Planning, Analysis, and Information Resources. The authors would like to thank Allen Schulenberg, Robert Kaufman, Kelvin LaCroix, Steve Gunning, Sam Pernici, Roth Aymond, and David Elder (Division Director), without whom this work would not have been possible.

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.

Table of Contents

Abstract

Value Added Assessment of Teacher Preparation

Analyses were conducted examining the feasibility of using Louisiana's student achievement, teacher, and curriculum databases to assess the efficacy of teacher preparation programs within Louisiana. Work began with the construction of a large multivariate longitudinal database linking many data points. This was followed by a model development phase in which mixed linear models were developed to predict student achievement based upon prior achievement, student demographic factors, and classroom level covariates. The model nested students within teachers and teachers within schools. The model included effects for teachers and schools. Separate models were developed for each content area. These models were used to assess the efficacy of teacher preparation programs. The same VAA model was applied to the educational data for the 2004-2005 and 2005-2006 school years. Examination of ELA indicated poor stability of estimates across years and raised issues regarding the alignments of the analysis with the way ELA teaching assignments are made. Additional work is planned examining the impact of assessing written expression and reading separately, which will more closely match the pattern of teaching assignments. Results also clarified the need to set a higher standard for the number of program completers necessary before reporting results. This issue appears to be addressable through the use of cross-year pooled analyses. The data also suggested that it is possible to identify teacher preparation programs that are relative outliers in their estimated contribution to student achievement. This is particularly encouraging in an analytic context in which students' prior achievement, students' demographic variables, classroom context variables, and school building level variables have all been included in the model. Additional analyses suggested that the bulk of the variance shared between a family demographic survey and student educational achievement was accounted for by data in Louisiana's educational databases. One limitation of the analyses reported herein is that the data for the bulk of the programs reflects graduates of the programs prior to their redesign in 2000-2003. These programs are being phased out and are no longer admitting students. The data for the few redesigned programs that have sufficient program completers was very encouraging, but is a very distinct part of the population of teacher preparation programs. As teacher preparation programs increasingly produce graduates from their redesigned programs it should be possible to assess the impact of an increasing number of redesigned programs and compare their results to their programs prior to redesign.

Interim Technical Report: Value Added Assessment of Teacher Preparation

I. Introduction

This technical report provides results for analyses completed as part of the *Value Added Assessment of Teacher Preparation Project* (VAA-TPP) housed in the Department of Psychology at Louisiana State University. The VAA-TPP is a multiyear study designed to examine the feasibility of using Louisiana's K-12 and higher education data systems to assess the impact of teacher preparation programs as pathways into teaching on student achievement. In this context the VAA-TPP is examining the impact of teacher programs as complete aggregates including recruitment of teacher candidates, admissions, content preparation, pedagogical preparation, field experiences, and selection for graduation. The VAA-TPP team does not have sufficient data to begin examination of these separate functions of teacher preparation programs. A separate statewide research team led by Dr. Jeanne Burns will be examining the processes of teacher preparation. Value added assessment as used in this context refers to the use of prior information (e.g., achievement and demographic data) to estimate expected outcomes (the level of student achievement) based on a large data system (all tested grades in Louisiana). This prediction of achievement is then used to assess the extent to which the performance of the students, within teachers' classrooms, nested within schools diverges from that prediction in a systematic manner. This information is used to estimate the degree to which new teachers' effectiveness is differentially associated with having entered teaching through particular teacher preparation programs (TPP).

The areas of value added assessment in education and educational production functions have a substantial and relatively rapidly growing literature base (see for example Ballou, Sanders, & Wright, 2004; Goldhaber & Brewer, 1997; Hill, Rowan, & Lowenberg, 2005; McCaffrey et al., 2003; Todd & Wolpin, 2003; Wayne & Youngs, 2003). Review of that substantial literature is beyond the scope of this interim technical report. This report provides a summary of the major findings of the analytic work completed in FY 2006-2007. Additional manuscripts will be prepared that discuss in greater detail the findings described herein, as well as additional analyses, in relationship to this rapidly developing literature base.

Prior Work

Initial pilot work was completed in 2004 and 2005 based on a sample of opportunity of 10 school districts in Louisiana in which it was possible to link teachers and students through administrative databases (Noell, 2004; Noell, 2005). These links in combination with achievement data, certification data, and demographic data were used to complete initial pilot work. The initial analyses suggested it was possible to detect some statistically significant differences for groups of new teachers that varied by content area and university preparation program. Additionally, the data suggested that the results were substantively robust to three different modeling specifications examined (Noell, 2004). Additionally, the estimates of TPP were strikingly similar across the two years for the two TPPs that provided enough data to be presented in analyses across years.

The initial analytic work using a statewide longitudinal database was completed and reported in 2006. The analytic approach adopted for those analyses was a hierarchical linear model (HLM; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002) that nested students within teachers and teachers within schools. HLMs have a number of desirable features for the analysis of educational data. First, they capture the natural nesting of students within higher level organizations such as classrooms and schools. Second, they permit correlation of error terms within nested units. This permits modeling of contextual and grouping variables in a statistically appropriate manner. Third, they provide a model in which units and their effects can appropriately be connected within a hierarchy (the effect of a school on teachers and through the teachers on students). Fourth, HLMs can provide estimates that are adjusted for observed unreliability of data (shrinkage estimates). Additionally, using a one year to the next school year approach within a covariate adjustment model, HLM results can be obtained that make relatively modest assumptions about the measurement qualities of the tests that are likely to be tenable (see Matrineau et al., 2007; Reckase, 2004; Seltzer, Frank, & Bryk, 1994).

Analyses that use only data from adjacent years lose the analytic power of examining multiyear achievement trajectories for students across multiple teachers (see McCaffrey et al., 2003; McCaffrey et al., 2004; Sanders & Horn, 1998; Todd & Wolpin, 2003), but avoid the difficult assumption that tests are repeated observations of a single unidimensional linear scale that is approximately invariant across grades (Matrineau, 2006; Seltzer et al., 1994) and result in a lower level of missing data issues. The single linear scale assumption is logically challenging for Louisiana's data given the shifting weighting of content assessed each year and the shifting nature of the content itself (see http://www.doe.state.la.us/lde/saa/2273.html for a description of the content by grade design of Louisiana's assessments). The degree to which this shift results in a shift in the constructs assessed or different multidimensional scales across near grade ranges (3 or 4 consecutive years) is an issue that can be examined in future research. Although a general language arts or mathematics construct hypothesis might be tenable for narrow grade ranges, it appears to be logically untenable for domains such as science and social studies where the content shifts so dramatically from year to year. It is also critical to note that at a very fundamental level, the shifting content of Louisiana's assessments is a strength of the assessments. Louisiana's assessments are aligned with the published grade level expectations for each grade. As a result, what is assessed is matched to the blueprint of what is to be taught.

The tradeoffs between analytic power, bias reduction, measurement assumptions, and missing data issues among the several viable approaches to fitting models in contexts that assess educational outcomes remains an active area of research and analysis (see for example Lockwood & McCaffrey, 2007; Martineau, 2006). However, the viability of models that assume an unchanging outcome construct across grades is problematic for some grades and subjects in Louisiana. Additionally, the loss of analyzable records due to grade retention is already a significant concern and is a problem that is substantially exacerbated by analytic models spanning additional years. 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 current analyses replicated the prior year approach (Noell, 2006) and adopted an HLM covariate approach to the data.

Context for the Current Analyses: Hurricanes Katrina, Rita, & New Tests

The analyses described herein are based primarily on the 2005-2006 school year in Louisiana. It is important to recognize that during that school year Hurricane Katrina came ashore in New Orleans, effectively destroying the largest school district in Louisiana. Additionally, Katrina severely affected the functioning of several neighboring school districts such as Plaquemines and St. Bernard. The magnitude of the disruption in a number of school districts is such that they provided no or virtually no records that could link students with achievement data to teachers. Additionally, varying levels of disruption are evident in the records for school districts radiating out from the point of landfall. For example, the Jefferson School District identifies a large percentage of its students as being hurricane affected while still providing a substantial number of complete usable educational records, but that number is substantially fewer than it was for 2004-2005. The decrease in available educational records that can contribute to the value added assessment appears to spiral out from the point of impact in varying degrees following the path of Katrina. For example, St. Tammany School District identified approximately 18% of its students who completed the end of year assessments (*LEAP* and *i*LEAP) as being hurricane disrupted. It provided 27% fewer student records eligible for analysis and 10% fewer teacher records. In parallel, the landfall of Hurricane Rita in Cameron Parish created substantial disruption in that area of the state, primarily in the Cameron School District. The Cameron School District provided 84% fewer usable teacher records and identified 96% of its students who completed the state end of year assessments as hurricane disrupted. The evidence suggesting disruption spiraling out from the Cameron School District is much less dramatic than that for Hurricane Katrina radiating out from New Orleans.

One of the original hopes of the authors was to combine selective deletion of school districts with the dropping of students identified as hurricane disrupted by the school districts to in part isolate the effects of the two natural disasters. Neither strategy appears to provide a compelling solution. The selected deletion strategy was problematic due to the progressively attenuating but diffuse nature of the hurricane effects. It is simply not entirely clear where lines should be drawn and any lines would likely drop large population centers that provided thousands of complete and analyzable records. The codes that the Louisiana Department of Education had the foresight to develop to identify students whose education was disrupted by hurricanes did not appear to identify students whose educational attainment was clearly adversely affected by the storms. The demographics of the storm affected areas resulted in data that closely reflect school district attendance. For example, approximately 84% of the public school students for whom a hurricane disruption code was reported for Hurricane Rita attended Calcasieu School District. Calcasieu had a net positive district level effect on student achievement for both 2004-05 and 2005-06 (Hurricane Rita) that in some models would paradoxically appear to be a positive effect for the impact of Hurricane Rita. A somewhat parallel issue arises in the New Orleans area in which more than half of the students whose education was identified as disrupted by Hurricane Katrina attended school in the Jefferson School District.

The essential underlying issue is one of large scale missing data. To a considerable extent, those students and school districts most strongly impacted by the hurricanes are simply not represented in the data. Although that may be desirable from some perspectives, it makes interpretations of data problematic for all teacher preparation programs to some degree and makes them especially difficult for others in relationship to the prior year's data. For all programs, the change in the teacher work force across the two years creates different norms for comparison to the experienced certified teachers. Hundreds of experienced certified teachers who formed a considerable part of the normative comparison group for all programs for the 2004-05 school year are not represented in the data system for 2005-06. Additionally, that attrition was highly selective and regional. The effects of this change are equally relevant to all teacher preparation programs (TPP) regardless of where they are located or where their graduates teach. Additionally, for some universities, but not others, selective attrition will have occurred in their graduates to the extent that they teach in Orleans, Cameron, St. Bernard, or Plaquemines School Districts in particular. This is particularly noteworthy given that the Orleans School District employed new teachers from a greater variety of TPPs than any other school district in Louisiana. In the year prior to Katrina, Orleans employed first through third year teachers from 13 different TPPs. That is more than half of the TPPs in mathematics represented in this report for 2004-05. The systematically missing data create challenges in comparing results to the prior academic year.

A final historical artifact that may impinge on these results is the shift in the testing protocol that occurred in AY 2005-2006. Louisiana shifted from the administration of the Iowa Test of Basic Sills (*ITBS*) to the *i*LEAP. The *i*LEAP includes a subset of *ITBS* items that make up the bulk of that scale, plus some augmentation for Louisiana. Generally the goal of the selective deletion and the augmentation is to bring the assessments administered more fully in line with the State's content standards as expressed in the published grade level expectations (GLEs). In the long term this should improve the content validity of the assessments by assuring tighter alignment between what is expected to be taught and what is assessed; however, any change in the assessment instruments has the potential for unintended consequences. This is particularly true in the initial year of implementation.

II. Data Merging Process

The 2005-2006 academic year was the primary target for initial analysis. Data analyses for 2004-2005 were also conducted to supplement the current year work and provide a point of comparison. The two school years were combined to obtain university estimates (described below). The data merging process for the 2004-2005 data are described in Noell (2006) and will not be repeated here. Data for 2005-2006 contributing to these analyses were drawn from the curriculum database linking students and teachers for 2005-2006 spring standardized testing assessments (*ITBS* and *LEAP-21*) for spring 2005 and spring 2006 (*i*LEAP *and LEAP-21*). Additional data drawn from student databases, teacher certification databases, and Board of Regents program completer databases were merged with the database. Initial work was undertaken to resolve duplicate records and multiple partially complete records that described the same student. 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. Those records that did not uniquely match at any stage were retained as isolated records of student performance. The matching process included five 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 records from the previous stage available at critical points in the database construction. Several important decision points are noteworthy. Initial records were limited to students who completed an assessment in grades 4-9 to permit the availability of one year prior achievement data (grade 3). A number of students were noted as only attending public school for part of the 2005-06 school year. Due to the difficulty of ascertaining who taught these students, they were dropped from further analyses. Additionally, a substantial percentage of students changed schools during the school year. Because the student-teacher-course nexus data are only collected once per year, once a student moves it is not possible to ascribe subsequent instruction to a particular teacher. Although future work may ultimately recommend assigning the students' achievement to a teacher in part, the reality that the currently available data result in the majority of instruction remaining unaccounted for for students who change schools resulted in the decision to drop these students in the current analyses. It is important to note that student mobility between schools with the 2005-06 school year was 35% higher than it had been for the 2004-05 school year based upon student attendance records. This is another indicator of the disruptive effects of the hurricanes that created substantially more missing data.

It is also important to note the impact of screening for attendance on how data relevant to hurricane affected areas will contribute to the analyses. In order for a student to be included in the analyses, the student and his/her teacher had to be linked at the beginning of the year and that student and teacher had to remain at the same school through the *entire* school year. This means that students and teachers who changed schools as a result of the hurricanes or any other factor were dropped from the analyses.

In most instances (93.2%), students for whom assessment data were available for 2006 were matched to assessment records for 2005. Students who were retained at the end of AY 2004-05 were also dropped because the meaning of assessment data for students who are repeating the same grade is different from students who were promoted. The percentage of students who were matched to their teachers in the curriculum database is expected to be attenuated within the 2005-06 school year due to the disruptive effects of the hurricanes in significant population centers in south Louisiana. Although the percentage of records lost at each decision point in most cases is not severe, the cumulative effect results in the loss of a large percentage of the records (approximately 36.6%). The bulk of the loss is due to mobility between schools within years. Given the limitations of the current data collection procedures in Louisiana, it does not appear likely that it will be possible to apportion these students' instruction among the many teachers who are likely to have taught them during that school year. Issues such as grade retention also do not appear to have a likely resolution in the near term. It is anticipated that in typical school years (lacking a natural disaster and not bordering one with a natural disaster), the percentage of complete records will increase considerably.

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

	English-Language Arts	**Mathematics**
Assessed students grades 4-9	299,811	297,226
In public school in Louisiana for the full year	282,304 (94.2%)	280,454 (94.4%)
Remained in one school for the full year	239,986 (85.0%)	238,772 (85.1%)
Matched to 2005 data	223,688 (93.2%)	222,526 (93.2%)
Promoted at the end of 2005	202,820 (90.6%)	201,867 (90.7%)
In curriculum database	189,916 (93.6%)	188,808 (93.5%)

Note. The percentage in parentheses within each cell is the percentage of records from the previous stage available in the current stage of database construction.

III. Preliminary Analyses

Prior to analyses that linked students and teachers within an HLM, a series of statewide ordinary least squares (OLS) regression analyses were conducted to examine general patterns in the data. Progressively more variables were employed as predictors and the multiple correlations between achievement in 2006 and predictor variables were examined. For purposes of these analyses, promotion was used as a screening factor, but mobility within the school year was not. Students who moved were included because these analyses did not attempt to link these students to their teachers. As a result more students were eligible to contribute to these analyses than for the analyses that link students and teachers described below. Test scores were standardized to a mean of zero and unit standard deviation within grade and year. Demographic and hurricane disruption variables (described below) were entered as dummy codes. First and second order polynomial terms for prior achievement were examined and not found to be statistically significant. Similarly, a large family of demographic interaction terms was examined, with prior achievement and demographic factors included in the equations, and were not found to be statistically significant. As a result, polynomial predictors for prior achievement in the content area assessed and demographic interaction terms are not presented below or discussed further.

It is worth noting that throughout this report, 2004-2005 results are reanalyzed and presented because of the availability of additional data that were not included in the FY 2005-2006 analyses. These include variables that were available, but had not yet been merged into the data system at LSU (detailed special education data) and data that had not yet been obtained from the Louisiana Department of Education (attendance data for students and teachers).

Variables were entered sequentially in blocks to examine the predictive power of conceptually meaningful blocks of variables: prior achievement, demographic factors, attendance, and reported natural disaster impact. Results for all four content areas are presented below followed by a brief discussion of the results.

Table 2*: English-Language Arts Statewide Regression Analyses for 2005 & 2006*

Predictors	Multiple correlation (Number of Students) 2005	Multiple correlation (Number of Students) 2006
Z-score Prior Year ELA	.759 (283,450)	.747 (253,308)
Z-scores Prior Year Achievement	.785 (282,518)	.789 (253,008)
Z-scores Prior Year Achievement Student demographic factors	.798 (282,515)	.810 (252,987)
Z-scores Prior Year Achievement Demographic & Hurricane data	*	.811 (252,987)
Z-scores Prior Year Achievement Demographic, Hurricane, and attendance	.801 (281,932)	.814 (252,582)
Z-score: Two Prior Year ELA	.809 (207,225)	.792 (189,044)
Z-scores: Two Prior Year Achievement	.819 (206,664)	.822 (188,595)
Z-scores Two Prior Year Achievement Student demographic factors	.825 (206,661)	.834 (188,580)
Z-scores Two Prior Year Achievement Demographic & Hurricane data	*	.834 (188,580)
Z-scores Two Prior Year Achievement Demographic, Hurricane, and attendance	.827 (206,304)	.836 (188,312)
Z-score: Three Prior Year ELA	.829 (152,223)	.798 (138,797)
Z-scores Three Prior Year Achievement	.837 (151,915)	.829 (138,388)
Z-scores Three Prior Year Achievement Student demographic factors	.841 (151,912)	.840 (138,381)
Z-scores Three Prior Year Achievement Demographic & Hurricane data	*	.840 (138,381)
Z-scores Three Prior Year Achievement Demographic, Hurricane, and attendance	.844 (151,696)	.842 (138,208)

Table Note. Prior Year achievement includes the *Z*-scores for ELA, mathematics, science, and social studies. *Student demographic factors* included were free lunch status, reduced price lunch, gifted status, primary special education diagnosis (codes for emotionally disturbed, specific learning disability, mild mental retardation, other health impaired, speech/language concerns, and other special education diagnosis), limited English proficiency status, gender, Section 504 eligibility, and minority status (codes for Asian American, African American, Hispanic, and Native American). Only a combined free/reduced lunch status variable was available for 2004-2005. *Hurricane Disrupted* included seven hurricane disruption codes provided by the Louisiana Department of Education. They provided separate codes for students disrupted from public schools, private schools, and out of state for Hurricane Katrina and Hurricane Rita plus an other hurricane disrupted code. *Attendance* was the number of days the student was absent.

Table 3: *Mathematics Statewide Regression Analyses for 2005 & 2006*

Predictors	Multiple Correlation (Number of Students) 2005	Multiple Correlation (Number of Students) 2006
Z-score Prior Year Math	.779 (283,296)	.772 (251,699)
Z-scores Prior Year Achievement	.798 (282,430)	.790 (251,459)
Z-scores Prior Year Achievement Student demographic factors	.806 (282,427)	.802 (251,438)
Z-scores Prior Year Achievement Demographic & Hurricane data	*	.803 (251,438)
Z-scores Prior Year Achievement Demographic, Hurricane, and attendance	.809 (281,844)	.805 (251,040)
Z-score: Two Prior Year Math	.823 (207,148)	.813 (187,629)
Z-scores: Two Prior Year Achievement	.830 (206,608)	.821 (187,235)
Z-scores Two Prior Year Achievement Student demographic factors	.833 (206,605)	.827 (187,220)
Z-scores Two Prior Year Achievement Demographic & Hurricane data	*	.828 (187,220)
Z-scores Two Prior Year Achievement Demographic, Hurricane, and attendance	.835 (206,248)	.830 (186,957)
Z-score: Three Prior Year Math	.841 (152,170)	.827 (137,534)
Z-scores Three Prior Year Achievement	.846 (151,875)	.832 (137,172)
Z-scores Three Prior Year Achievement Student demographic factors	.847 (151,872)	.837 (137,165)
Z-scores Three Prior Year Achievement Demographic & Hurricane data	*	.837 (137,165)
Z-scores Three Prior Year Achievement Demographic, Hurricane, and attendance	.850 (151,656)	.840 (136,996)

Table Note. All variables were entered as in Table 2, see the note above.

Table 4*: Science Statewide Regression Analyses for 2005 & 2006*

Predictors	Multiple correlation (Number of Students) 2005	Multiple correlation (Number of Students) 2006
Z-score Prior Year Science	.722 (282,113)	.690 (211,544)
Z-scores Prior Year Achievement	.768 (282,072)	.747 (211,528)
Z-scores Prior Year Achievement Student demographic factors	.779 (282,069)	.761 (211,513)
Z-scores Prior Year Achievement Demographic & Hurricane data	*	.761 (211,513)
Z-scores Prior Year Achievement Demographic, Hurricane, and attendance	.780 (281,493)	.764 (211,196)
Z-score: Two Prior Year Science	.776 (206,418)	.737 (150,442)
Z-scores: Two Prior Year Achievement	.800 (206,397)	.772 (150,427)
Z-scores Two Prior Year Achievement Student demographic factors	.804 (206,394)	.779 (150,416)
Z-scores Two Prior Year Achievement Demographic & Hurricane data	*	.779 (150,416)
Z-scores Two Prior Year Achievement Demographic, Hurricane, and attendance	.805 (206,040)	.781 (150,217)
Z-score: Three Prior Year Science	.800 (151,743)	.756 (103,835)
Z-scores Three Prior Year Achievement	.819 (151,716)	.784 (103,831)
Z-scores Three Prior Year Achievement Student demographic factors	.822 (151,713)	.789 (103,826)
Z-scores Three Prior Year Achievement Demographic & Hurricane data	*	.789 (103,826)
Z-scores Three Prior Year Achievement Demographic, Hurricane, and attendance	.823 (151,498)	.791 (103,705)

Table Note. All variables were entered as in Table 2, see the note above.

Table 5*: Social Studies Statewide Regression Analyses for 2005 & 2006*

Predictors	Multiple Correlation (Number of Students) 2005	Multiple Correlation (Number of Students) 2006
Z-score: Prior Year Social Studies	.686 (281,999)	.621 (200,364)
Z-scores Prior Year Achievement	.745 (281,990)	.693 (211,366)
Z-scores Prior Year Achievement Student demographic factors	.751 (281,987)	.705 (211,351)
Z-scores Prior Year Achievement Demographic & Hurricane data	*	.705 (211,351)
Z-scores Prior Year Achievement Demographic, Hurricane, and attendance	.753 (281,411)	.709 (211,033)
Z-score: Two Prior Years Social Studies	.738 (206,361)	.660 (150,310)
Z-scores Two Prior Years Achievement	.772 (206,256)	.701 (150,304)
Z-scores Two Prior Years Achievement Student demographic factors	.774 (206,353)	.709 (150,293)
Z-scores Two Prior Years Achievement Demographic & Hurricane data	*	.709 (150,293)
Z-scores Two Prior Years Achievement Demographic, Hurricane, and attendance	.775 (205,999)	.712 (150,094)
Z-score: Three Prior Years Social Studies	.763 (151,688)	.689 (103,745)
Z-scores Three Prior Years Achievement	.789 (151,682)	.721 (103,743)
Z-scores Three Prior Years Achievement Student demographic factors	.790 (151,679	.727 (103,738)
Z-scores Three Prior Years Achievement Demographic & Hurricane data	*	.727 (103,743)
Z-scores Three Prior Years Achievement Demographic, Hurricane, and attendance	.791 (151,464)	.731 (103,617)

Table Note. All variables were entered as in Table 2, see the note above.

Across all content areas and both academic years, prior achievement in each content area was strongly related to current year achievement. The relationship was weaker for social studies and science. The combination of the four prior achievement scores for the prior year was a strong predictor of current year achievement (multiple R = .79) for both ELA and mathematics. Interestingly, across all four content areas the increment in the multiple r when entering a block of 16 demographic variables with prior year achievement already in the model was small and fairly consistent (range .007 to .015). In ELA and mathematics the increased multiple R when demographic variables were included resulted in an increase in shared variance of approximately 2% (mathematics) to 3% (ELA). Adding another year of achievement data in ELA and mathematics resulted in a modest additional increment in the multiple R (+.022 to +.025). It is interesting to note that the relationships between years' testing data was reasonably consistent for ELA and mathematics, but appeared to change for science and social studies to a degree that may not reflect chance fluctuations. The relationship between prior test performance and current year performance decreased in both cases, but more markedly for social studies. One prior year's achievement data combined with demographic variables and attendance demonstrated a strong relationship with current year achievement for mathematics and ELA (multiple $r > .8$). The relationship was somewhat weaker for science and weaker still for social studies.

IV. Linking Students and Teachers

Following preliminary linking of data and student level analyses, the student achievement data were linked with the data connecting students to courses and courses to teachers. In addition, data from the Profile of Educational Personnel (PEP) and the certification database provided by the Louisiana Department of Education's Division of Planning, Analysis, and Information Resources were linked to teachers and the longitudinal educational achievement database. Data provided by the Louisiana Board of Regents regarding teacher preparation program completers were used to identify new teachers.

Course codes were collapsed into groups that were associated with specific test areas (i.e., ELA, mathematics, science, and social studies). For example, English I was associated with ELA tests and Life Science with science tests. 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. So 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.

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

Replicating the general approach from Noell (2006), the educational assessment data were analyzed using hierarchical linear models (HLM; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002). Three hierarchical structures were examined: students nested within classrooms, students within classrooms with schools, and students within classrooms within school districts. Table 6 presents the findings for ELA and mathematics for teachers, schools, and school districts in the three different nesting hierarchies. These models included prior achievement and student demographics as predictors of student achievement.

Table 6: *Preliminary Shared Variance for Different Nesting Hierarchies that include Student and Classroom Level Predictors*

		Model		
Content	Variance Component	Students within Teachers	Students within Teachers within Schools	Students within Teachers within School Districts
Mathematics	Teachers	13.1%	10.0%	12.3%
	Schools		3.3%	
	School Districts			1.0%
	Students	86.9%	86.7%	86.7%
ELA	Teachers	13.3%	10.1%	13.2%
	Schools		3.7%	
	School Districts			0.8%
	Students	86.7%	86.2%	86.0%

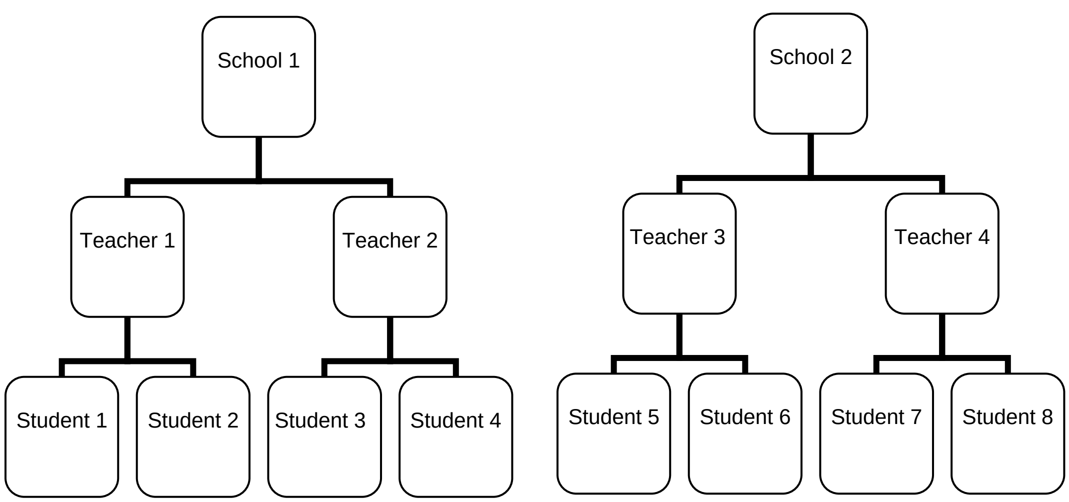
Overall, the proportions of variance associated with each level within each of the three nesting structures are extremely similar to the findings for 2004-2005 (Noell, 2006). Notably, the bulk of the variance in current year achievement lies between students. Given that students in grades 4 through 9 have considerable educational and developmental histories it makes sense that the bulk of current year test performance will be determined by student level factors such as individual differences, developmental history, and prior educational experience. It is important to recognize that the single year covariate adjustment approach used herein attempts to estimate the contribution of current year schools and teachers to total achievement as assessed by a cumulative examination. As a result, the variance estimates will necessarily be smaller than studies that have access to vertically aligned tests and can measure growth within the current year. An assessment model that removes prior years' achievement and isolates only current year learning would result in a much larger estimate of the contribution of current year teachers for conceptual and pragmatic reasons (see McCaffrey et al., 2003 and Rowan et al., 2002). However, Louisiana's tests are not vertically scaled and as a result, an analysis that attempts to isolate current year learning from total achievement requires making problematic assumptions regarding the nature of the testing inputs (Matrineau et al., 2007; Seltzer et al., 1994). The common subtraction of a prior year score from the current score, if advised at all, would not reflect growth. It would reflect movement within the distribution of achievement scores and it is not clear that this is a construct that would be widely and readily interpretable. It is also unclear whether this measure would be more informative to policy makers.

The dominance of individual differences notwithstanding, a non-trivial portion (approximately 13%) of the variance associated with current year test performance is associated with factors other than the student. Replicating the findings from 2004-2005, the data suggest the utility of a hierarchical structure that includes students within classes within schools. The proportion of the variance at the school building level is 3 to 4 times that at the school district level, suggesting the use of schools as a third level of the hierarchy over school districts. Additionally, of the variance that is not associated with students, approximately 26% of that appears to be between schools and 74% between teachers, suggesting that schools are an important factor beyond simply being a collection of teachers.

It is interesting to note the relationship between the estimated variance between teachers and schools versus students relative to the duration of students' educational careers. For the students contributing to this analysis, the school year being studied is on average 1 year out of an average 7.5 years of total school attendance or 13% of their school career to date. This figure corresponds closely to the proportion of variance the data suggest lies between teachers/schools in this one year. It is important to note that although these data suggest that the typical effect of one teacher in one year to total achievement to date in grade 4 to 9 does not appear to be large, that over several years the cumulative effects of teachers would be expected to be substantial (e.g. Sanders & Horn, 1998).

Figure 1 below depicts the final nesting structure that was employed.

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



Building the current models. The modeling approach was somewhat parallel to Tekwe and colleagues (2004) in general strategy and followed Noell (2006). The approach was replicated across ELA, mathematics, 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. Next, the students' prior year's achievement in ELA, mathematics, science, and social studies were entered as a block as fixed effects. All effects were significant in all content areas and were retained. Next, the 16 demographic variables employed in the regression analyses described above, student absences, and the hurricane disruption variables for Hurricanes Rita and Katrina were entered as a block. The demographic variables are presented in Table 7 along with the percentages of students for whom the demographic variable was coded as true among the students who were eligible to contribute to the regression analyses. Variables were then removed one at a time in order of the lowest *t* value until all remaining effects were significant.

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 will typically be strongly bounded by parental intervention. This is not so much an issue of absolute contributions but relative contribution to student absence. The authors adopted the assumption that students' absences were 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.

Table 7: *Student Level Demographic Variables*

Variable	**Percentage of Sample**
Gender (Male)	51.7%
African American	45.4%
Hispanic	2.1%
Asian American	1.2%
Native American	0.8%
Receiving Free Lunch	56.3%
Reduced Lunch	7.8%
Gifted	4.7%
Special Education: Emotionally Disturbed	0.8%
Special Education: Learning Disability	6.7%
Special Education: Mild Mental Retardation	1.1%
Special Education: Other Health Impaired	1.9%
Special Education: Speech Language	1.7%
Special Education: Other	1.7%
Section 504 Identification	5.1%
Limited English Proficiency	1.1%
Disrupted Education: Hurricane Katrina	11.6%
Disrupted Education: Hurricane Rita	5.3%

Once a model for student level achievement was developed, several classroom variables were examined. These variables were entered at the teacher/classroom level and were conceptualized as contextual factors that may moderate student achievement in addition to teachers. The variables that were examined are presented in Table 8.

Table 8: *Classroom 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
Class mean prior achievement in ELA
Class mean prior achievement in mathematics
Class mean prior achievement in science
Class mean prior achievement in social studies
Teacher absences
Percentage of students reported disrupted by Hurricane Katrina
Percentage of students reported disrupted by Hurricane Rita

As with the student level demographic factors these classroom variables were entered as a block and removed one at the 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. Classroom level variables accounted for a modest portion of the variance in student achievement. For example, entry of covariates at the classroom level of the model accounted for 1.4% of the total variance in student achievement in ELA and 7.5% of the variance at the teacher/classroom level. In mathematics, entry of classroom covariates reduced the variance component at the teacher level by 5.6%. These data suggest that classroom composition accounts for a modest portion of the variability in teacher effects, once detailed information about the students individually is already included. These compositional effects would certainly be much larger if student level data were not already included.

The same modeling process was then implemented across content areas for level 3 of the model (schools). The variables that were initially entered as a block are listed in Table 9. The few variables retained at the school level accounted for a very small proportion of the total variability in achievement after accounting for student level and classroom variables. For example, school building level covariates accounted for only 4.6% of the variance between schools in ELA. This represents an exceedingly small portion of the total variability (approximately 0.2%) when one considers that only 3.7% of the variance in achievement was between schools as opposed to students and teachers.

Table 9: *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
Class mean prior achievement in ELA
Class mean prior achievement in mathematics
Class mean prior achievement in science
Class mean prior achievement in social studies
Percentage of students reported disrupted by hurricane

The following tables 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, attendance, and hurricane variables that were retained were entered 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 potential equations that are not presented (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.

Table 10: *Hierarchical Linear Model for ELA Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Prior year ELA test	17.6	(17.3, 17.9)
	Prior year Math test	5.9	(5.6, 6.1)
	Prior year Science test	7.4	(7.1, 7.7)
	Prior year Social Studies test	7.7	(7.3, 8)
	Emotionally Disturbed	-16.8	(-21.4, -12.1)
	Section 504	-8.4	(-9.5, -7.3)
	Mild Mental Retardation	-41.0	(-46.5, -35.5)
	Other Health Impaired	-14.8	(-16.6, -13)
	Speech and Language	-5.1	(-6.3, -4)
	Specific Learning Disability	-23.6	(-25.2, -22)
	Special Education - Other	-6.3	(-7.8, -4.8)
	Gifted	9.2	(8.3, 10.2)
	Free lunch	-2.9	(-3.3, -2.5)
	Reduced price lunch	-1.4	(-1.9, -0.9)
	Gender (male)	-7.8	(-8.2, -7.4)
	Asian American	3.3	(1.7, 4.9)
	African American	-1.6	(-2.1, -1.2)
	Student Absences	-0.33	(-0.36, -0.30)
Classroom variables	% Special Education	-1.2	(-1.5, -.9)
	% Limited English Proficiency	-1.0	(-1.8, -.2)
	% Free Lunch	-0.7	(-1.1, -.4)
	Teacher Absences	-0.04	(-0.06, -0.02)
Building Variables	% Free lunch	0.6	(0.2, 1.1)
	Mean prior year ELA test	5.8	(3.4, 8.2)

The coefficients are scaled to the approximate standard deviation of the educational assessments (*i*LEAP and *LEAP*) used in Louisiana: 50. So after considering all other variables, a student who was Emotionally Disturbed would be predicted to score 16.8 points lower than one who was not and a student who was gifted would be predicted to score 9.2 points higher in ELA.

It is also 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 universities 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.

It is important to note that differences in how variables were scaled create the need for 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.

The largest single contributor to a student's ELA achievement *among the achievement predictors* was his or her achievement in that domain the prior year. The coefficient for prior achievement in ELA was more than twice the value of any other prior achievement variable's coefficient.

Among the demographic variables, the special education disabilities Mild Mental Retardation, Learning Disability, Emotionally Disturbed, and Other Health Impaired had large negative coefficients. In contrast, giftedness had a large positive coefficient above and beyond those students' generally high prior scores. The coefficients for the two ethnicity based demographic variables were relatively small and it is worth noting that the effects for Hispanic American and Native American were not statistically significant.

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. So a student who was absent 20 days would be predicted to score 6 points lower than one with perfect attendance.

Classroom demographic variables loaded in what would be the commonly expected direction. Classrooms with a high percentage of special education students, limited English proficiency students, and students receiving free lunch would be expected to score more poorly than one that was not similarly disadvantaged. For example, in a classroom in which 60% of the students received free lunch, the predicted achievement for each student would be 3.5 points lower than for a classroom in which only 10% of the students received free lunch (classroom demographic effects were scaled to 10% units of demographic change).

The school building coefficient for each 10% of the student body receiving free lunch was fairly modest and was positive. This may have provided a corrective factor for the combined contributions of free lunch at both the student and classroom level or may have served as a marker for the provision of special services (e.g., Reading First); in higher poverty schools. The effect for prior achievement in ELA seems readily interpretable as reflecting the cumulative disparity so often observed in schools. Students attending schools with higher prior aggregate achievement are predicted to perform better on current year assessments.

Table 11: *Hierarchical Linear Model for Mathematics Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Prior year ELA test	5.9	(5.6,6.1)
	Prior year Math test	25.0	(24.7,25.4)
	Prior year Science test	4.5	(4.2,4.7)
	Prior year Social Studies test	2.2	(2,2.5)
	Emotionally Disturbed	-10.1	(-14.1,-6.1)
	Section 504	-5.4	(-6.3,-4.5)
	Mild Mental Retardation	-25.1	(-30.3,-20)
	Other Health Impaired	-11.6	(-13,-10.2)
	Speech and Language	-3.0	(-4.1,-1.9)
	Specific Learning Disability	-12.5	(-13.7,-11.3)
	Special Education - Other	-13.3	(-14.8,-11.8)
	Gifted	13.0	(11.9,14.1)
	Gender (Male)	2.9	(2.5,3.2)
	Free Lunch	-1.6	(-1.9,-1.2)
	Asian American	7.6	(6.2,8.9)
	African American	-6.0	(-6.4,-5.5)
	Student Absences	-0.33	(-0.31,-0.35)
	Disrupted due to Hurricane Katrina	-4.2	(-5.7,-2.6)
Classroom variables	% Special Education	-0.9	(-1.2,-0.6)
	% Gifted	1.0	(0.7,1.4)
	% Free lunch	-0.8	(-1.0,-0.5)
	Teacher absences	-0.06	(-0.09,-0.03)
	Mean prior year achievement in ELA	2.4	(0.5,4.2)
	Mean prior year achievement in Math	-5.1	(-7.2,-2.9)
Building variables	% Section 504	1.1	(0.2,2.1)
	Mean prior achievement in Math	11.2	(8,14.4)
	Mean prior achievement in Science	-7.9	(-10.9,-4.8)

As noted above it is critically important to bear in mind the differences in scaling and the meaning of those scales if one attempts to compare coefficients across different types of predictors.

The largest single contributor to the prediction of 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 four times the value of any other prior achievement variable's coefficient. Similar to ELA, large negative effects were associated with specific special education diagnoses and a substantial positive effect was associated with being diagnosed as gifted. In contrast to ELA, the two ethnicity demographic codes that were retained had more moderate coefficients, while the effects for Hispanic American and Native American were found not to be statistically significant again. The magnitude of the coefficient for student absences is the same as the coefficient for ELA to two decimal places.

The contributions of classroom demographic variables to the prediction are similar in magnitude to those for ELA and generally in the expected direction. Students attending classes with more advantaged students (i.e., gifted) and fewer disadvantaged students (i.e., special education and receiving free lunch) would be expected to perform better. Similarly, students in classes with peers who performed well on the ELA test the year before would be expected to perform better, if only modestly. The result for mathematics appears paradoxical in that students who are in classes with peers who performed better would be expected to perform more poorly on the mathematics test this year. Although the meaning of this finding is somewhat uncertain, it would appear that this may serve as a corrective loading to attenuate the strong positive loadings for prior mathematics achievement at both the student and school levels.

The school building coefficient for each 10% of the student body identified as receiving Section 504 services was modest and positive. This may have served as a marker for the provision of more special services in these schools. The effect for prior achievement in mathematics seems readily interpretable. Students attending schools with higher prior aggregate achievement in mathematics are predicted perform better on current year assessments. The negative loading for science achievement is difficult to interpret as other than a capitalization on a small fluctuation in a very large dataset or a corrective loading to adjust for excessive cumulative positive prediction contributed by other variables in aggregate. It is worth noting that at the student level, school mean prior achievement in science was positively correlated with current year mathematics achievement ($r = .37$) in the absence of other predictors.

Table 12: *Hierarchical Linear Model for Science Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Prior year ELA test	5.9	(5.6, 6.2)
	Prior year Math test	8.2	(7.9, 8.5)
	Prior year Science test	13.4	(13.1, 13.8)
	Prior year Social Studies test	9.3	(9, 9.6)
	Emotionally Disturbed	-10.4	(-14.6, -6.1)
	Mild Mental Retardation	-26.6	(-31, -22.3)
	Other Health Impaired	-9.5	(-11.1, -7.8)
	Speech and Language	-3.5	(-4.7, -2.3)
	Specific Learning Disability	-13.4	(-14.6, -12.2)
	Special Education - Other	-5.7	(-7.2, -4.2)
	Gifted	6.2	(5.3, 7.1)
	Section 504	-4.1	(-5.1, -3.1)
	Limited English Proficiency	-3.0	(-5.5, -0.5)
	Free lunch	-2.3	(-2.7, -1.9)
	Gender (male)	4.9	(4.6, 5.3)
	Hispanic American	-1.5	(-2.8, -0.2)
	African American	-8.3	(-8.9, -7.7)
	Student Absences	-0.27	(-0.24, -0.30)
Classroom variables	% Special Education	-1.1	(-1.5, -0.8)
	% Gifted	0.9	(0.5, 1.3)
	% Free Lunch	-0.8	(-1.0, -0.5)
	% Minority	-0.1	(-0.3, 0.1)
	Teacher Absences	-0.06	(-0.09, -0.03)
	Mean prior achievement in Social Studies	-5.2	(-6.9, -3.6)
Building Variables	% Gifted	-1.2	(-2.0, -0.3)
	Mean prior achievement in Science	8.0	(5.6, 10.3)

As noted above it is critically important to bear in mind the differences in scaling and the meaning of those scales if one attempts to compare coefficients across different types of predictors.

For science the differentiation between the content areas was substantially more modest than for ELA or mathematics. Although prior science score was the strongest predictor, prior scores in mathematics and social studies were not terribly dissimilar. The contribution of the special education diagnoses, limited English proficiency, and free lunch were also similar to ELA and mathematics, but the magnitudes for the special education categories were somewhat smaller. Interestingly, being male was positively related to science achievement at a level that is not likely to be trivial on a state level scale. Unfortunately, being African American was negatively related to science achievement at a level that is very likely to have social significance at a statewide scale and is similar to the adverse loading for being African American for mathematics. The magnitude of the coefficient for student absences is very similar to the coefficients for ELA and mathematics.

The contributions of classroom non-achievement demographic variables to the prediction are similar in magnitude and direction to those for ELA and mathematics. Students attending classes with more advantaged students (i.e., gifted) and fewer disadvantaged students (i.e., special education and receiving free lunch) would be expected to perform better. Interestingly, the same apparent paradoxical result was obtained for science as for mathematics in which aggregate classroom achievement loaded negatively. Again, it is important to recognize that this is a phenomenon that only occurs in the context of a system in which a tremendous amount of information is already available regarding student achievement.

The school building level coefficient for each 10% of the student body identified as gifted was modest and negative. Similar to mathematics, mean school building level prior achievement was positively related to subsequent achievement.

Table 13: *Hierarchical Linear Model for Social Studies Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Prior year ELA test	6.5	(6.2, 6.9)
	Prior year Math test	6.0	(5.7, 6.4)
	Prior year Science test	9.9	(9.6, 10.2)
	Prior year Social Studies test	12.0	(11.7, 12.3)
	Emotionally Disturbed	-8.4	(-12.4, -4.4)
	Section 504	-6.3	(-7.4, -5.2)
	Mild Mental Retardation	-26.0	(-30.8, -21.3)
	Other Health Impaired	-11.9	(-13.8, -10)
	Speech and Language	-2.4	(-3.8, -1)
	Specific Learning Disability	-14.6	(-15.9, -13.3)
	Special Education - Other	-2.8	(-4.5, -1.1)
	Gifted	3.6	(2.6, 4.5)
	Gender (male)	3.7	(3.3, 4.1)
	Free lunch	-3.3	(-3.7, -2.8)
	Reduced price lunch	-1.6	(-2.2, -1)
	Hispanic American	3.2	(1.9, 4.6)
	Asian American	5.4	(3.9, 7)
	African American	-3.6	(-4.1, -3)
	Student Absences	-0.33	(-0.31, -0.33)
	Disrupted due to Hurricane Rita	1.6	(-0.2, 3.4)
Classroom variables	% Minority	-0.3	(-0.5, 0.0)
	% Special Education	-0.7	(-1.1, -0.3)
	% Free Lunch	-0.4	(-0.8, -0.1)
	Teacher Absences	-0.07	(-0.1, -0.04)
	Mean prior achievement in ELA	3.8	(1.8, 5.8)
	Mean prior achievement in Social Studies	-5.3	(-7.5, -3)
Building variables	% Section 504	0.1	(-0.03, 0.2)
	Mean prior achievement in Science	8.6	(6.1, 11.1)

The results for social studies are similar to science in the more modest degree of differentiation of the content areas as predictors of social studies achievement. The magnitudes of the coefficients for special education disability status are similar to those for mathematics and science. Interestingly the coefficient for giftedness was considerably smaller than it was for the other content areas. Positive statistically significant coefficients were obtained for Hispanic and Asian Americans, while the coefficient for African Americans was negative. The magnitude of the coefficient for student absences was similar to the other content areas.

The results for the non-achievement demographic variables at the classroom level for social studies was similar to the other content areas, but with a somewhat smaller magnitude. Similar to mathematics, mean prior achievement in social studies loaded negatively in the model. It is interesting to note that a paradoxical loading (relative to common expectations) occurred for three of the four content areas for prior achievement at the classroom level. Although it was statistically significant, the loading for students identified as protected by Section 504 is sufficiently small that it is unlikely to be of much consequence. The strong positive loading for science at the school building level is similar to the other content areas except that it is a different content that is loading at the school level rather than the targeted content.

Summary. Generally the student level models were quite similar. For all areas, prior achievement in the target content area had the largest coefficient *among prior achievement variables*, with achievement in the other three content areas loading to varying degrees. Having a special education diagnosis was a consistent strong negative predictor of achievement with this being especially so for students identified as Mildly Mentally Retarded, Emotionally Disturbed, Other Health Impaired, and Learning Disabled. Student absences and free lunch status exhibited consistent relatively small coefficients. Among the ethnicity factors, the code for Asian American exhibited a positive coefficient across all content areas and African American a negative coefficient. The magnitudes varied across content areas. Gender, coded as male, was the only student level variable for which the direction of the coefficient changed across content areas. It was negative for ELA and positive for mathematics, science, and social studies.

Generally, the coefficients for non-achievement demographic variables were consistent at the classroom level with small positive coefficients for increasing percentages of advantaged groups (e.g., gifted students) and negative coefficients for disadvantaged students (e.g., disabled students). Teacher absences exhibited a small negative effect with approximately 20 teacher absences resulting in a decreased prediction for student achievement of 1.2 points across content areas. In three of four content areas a counter intuitive loading occurred in which higher classroom mean prior achievement resulted in lower predicted outcomes. However, this is a phenomenon that only emerges in a model that is saturated with a tremendous amount of information about student achievement and demographic factors. The simple relationships are in the expected directions.

No particularly clear pattern of variables emerged at the school building level other than the consistent positive coefficient for mean prior achievement at the school building level in the target subject (3 of 4 cases) or another subject (social studies).

VI. Assignment of Teachers to Groups

The operational definition of “new” teacher adopted in the early pilot work by this research team had been that teachers in their first three years of teaching were new (Noell, 2004; Noell, 2005). However, review of data for the 2004-2005 school year suggested that at the third year of teaching, teacher effectiveness was reasonably comparable to experienced teachers. The examination of years’ experience effects was replicated in the current year with a modified methodology and extended to science and social studies. The present analyses were conducted using dummy codes for each year of experience from 0 to 20. Teachers with 21 to 30 years experience were used as the comparative group. This 10 year band of veteran teachers was used as the comparative anchor because it was anticipated to provide a large stable base against which to compare and it avoided potential instability introduced by selecting one centrally located cohort (e.g., teachers with 15 years of experience) as the comparative point. This method for describing the change in teacher effectiveness across experience cohorts was adopted to avoid the issues of shrinkage introduced by the prior residual based method. The same models that were developed for the 2005-2006 data were also fitted to the 2004-2005 data to obtain an additional estimate of the effect of years experience cohort membership. The estimates for ELA, mathematics, science, and social studies are presented in Figures 2 through 5 below.

Figure 2

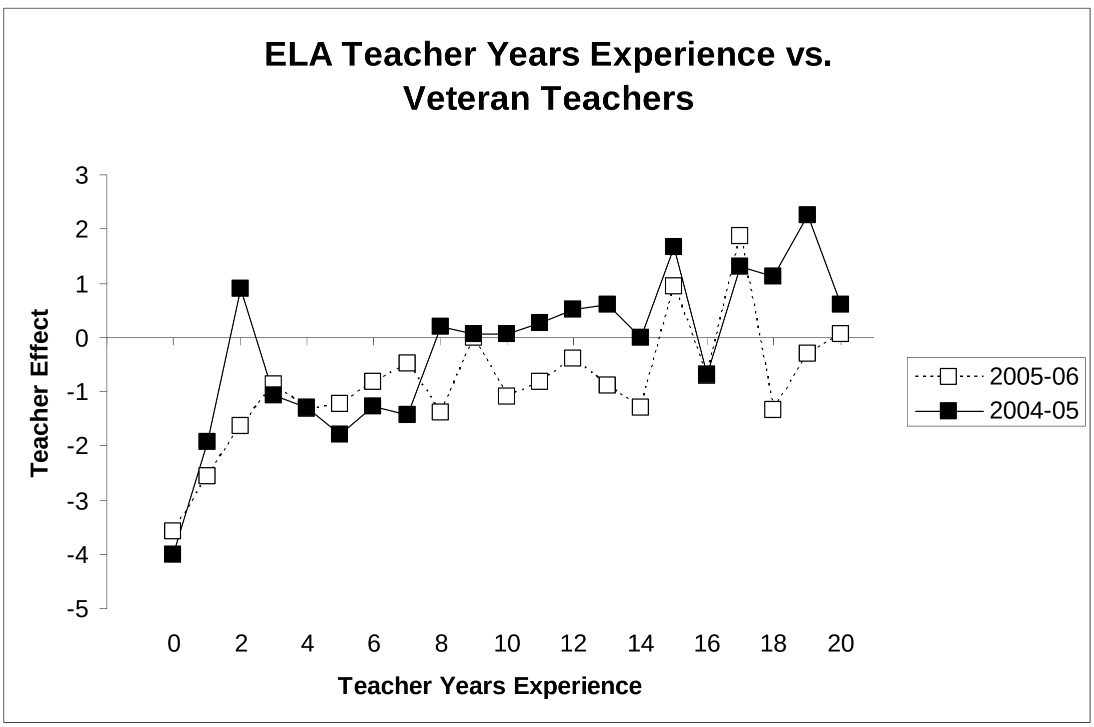


Figure note. The zero line on this graph represents the average effect for teachers with 21-30 years experience.

Figure 3

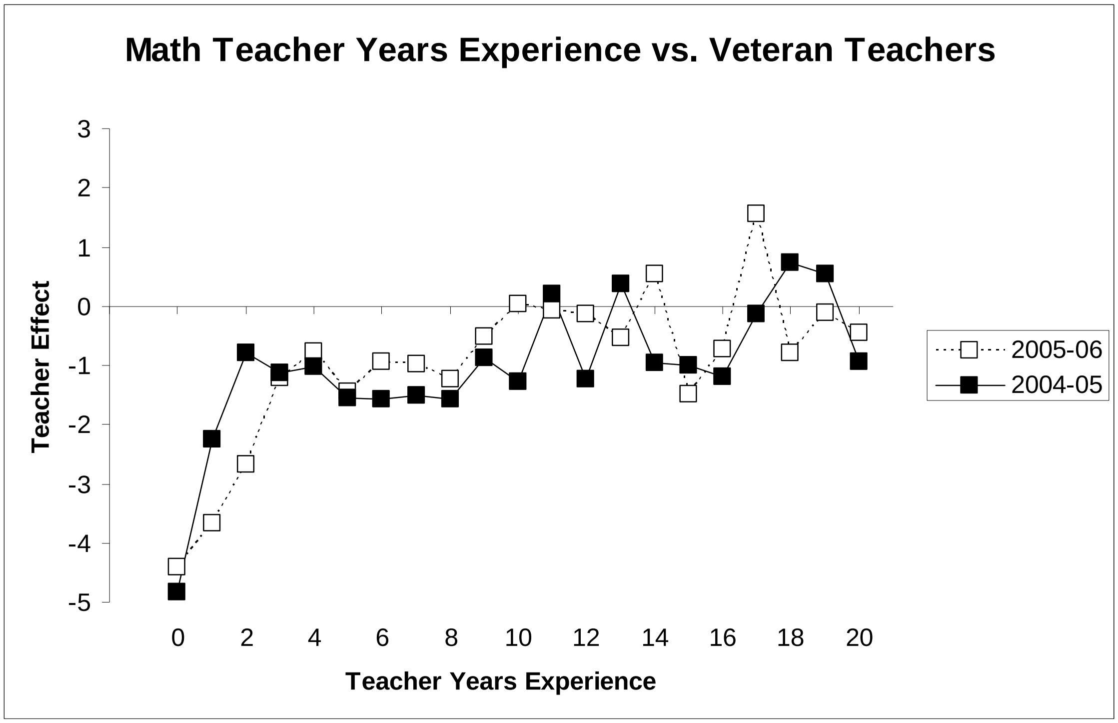


Figure note. The zero line on this graph represents the average effect for teachers with 21-30 years experience.

Figure 4

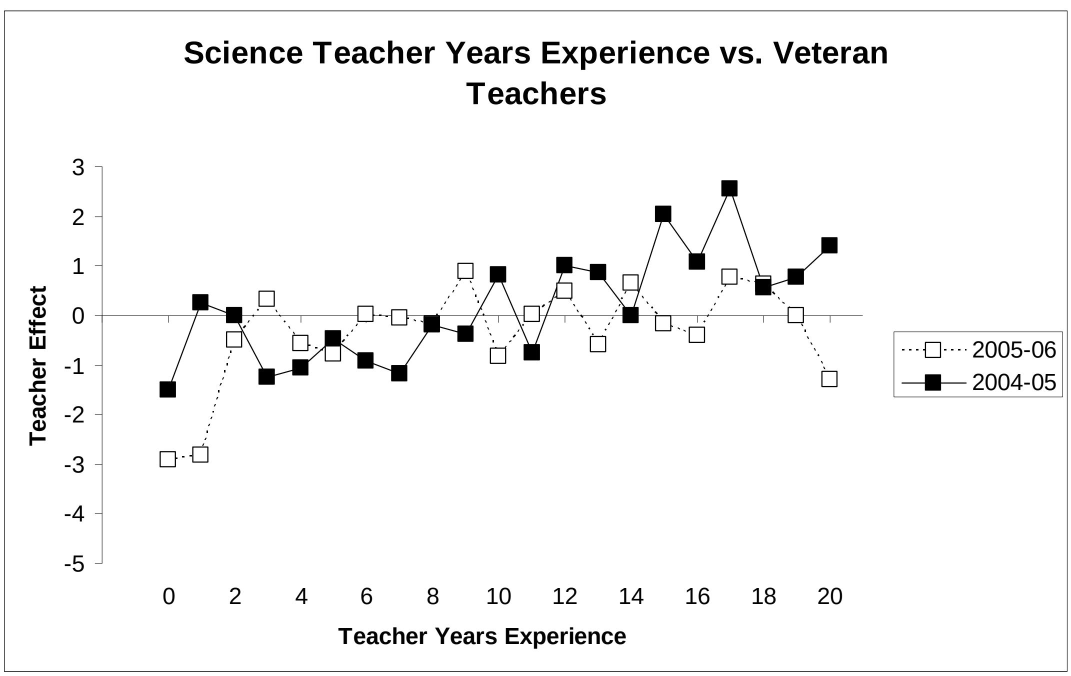


Figure note. The zero line on this graph represents the average effect for teachers with 21-30 years experience.

Figure 5

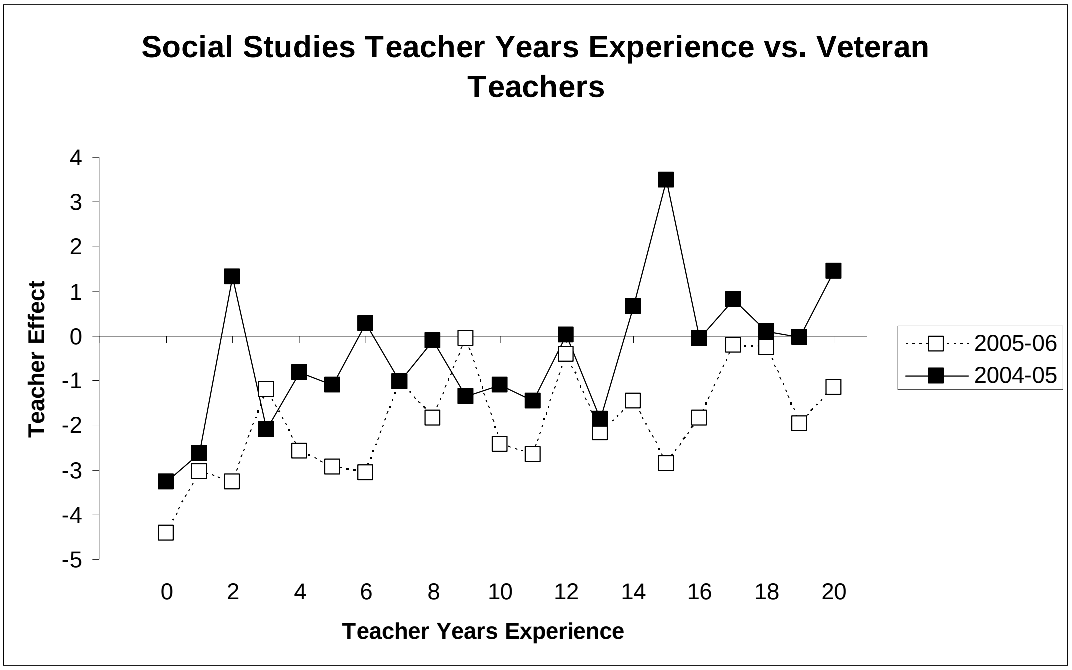


Figure note. The zero line on this graph represents the average effect for teachers with 21-30 years experience.

Across years assessed and content areas similar, but not entirely consistent, pictures emerged of the changes in teacher effects by experience cohorts. In ELA the 2005-2006 data suggest identifying teachers in their first three years of teaching as new and for 2004-2005 the data would suggest the first two years. In both cases, teacher cohorts become more like experienced teachers with each successive year. This same pattern is evident in mathematics. In science, for 2004-2005 a weak argument could be made for first year teachers being less effective; the data suggest no clear new teacher period. First and second year teachers were distinct from more experienced teachers for 2005-2006 in science.

The data for social studies is less orderly with a dip in estimated teacher effectiveness for the period from four to six years that occurs only in 2005-2006 and a generally lower level of estimated teacher effectiveness relative to veteran teachers than for 2004-2005. The reason for this difference is not clear. This may be an instance in which the combination of the hurricanes and/or the changes in the test may have created a transitory phenomenon in the data that will not be well understood. It is also important to note that beginning in 2005-2006, students in 9th grade are no longer being assessed in science and social studies. As a result 9th grade students for 2004-2005 were excluded from this analysis and the subsequent VAA for science and social studies. For social studies 2004-2005 some evidence for a weaker performance in the first two years of teaching are evident. For 2005-2006 the dip in performance for 4 to 6 years experience makes interpretation of the data less clear.

Collectively, the data created plausible arguments for treating both first through second and first through third year teachers as new. The decision was made to retain first and second year teachers as the target group for analysis based on the findings that two years was supported more often than three years. Additionally, the data that were not confounded by the hurricanes and adoption of new tests (2004-2005) were given a somewhat greater weight in the decision.

It should also be noted that the research team has learned that the meaning of the years experience variable is not entirely consistently implemented across school districts. Since that variable is tied to educator pay, some differences in decisions across school districts lead to error variance in the data. For example, in one school district, new teachers may be given experience credit for years spent working with youth outside schools as a recruiting tool, while in other districts only years of classroom experience in public schools may be awarded teaching experience. Future work is planned using the Department of Education's employment databases to create a new experience variable that may be less subject to this confound and structure analyses that capitalize on the repeated observations across teachers once this more accurate database is available.

Table 14: *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 teaching 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' degree granting institution for a bachelor's degree that could lead to teacher certification or completion of an alternative teacher certification program.

VII. VAA of Teacher Preparation

Once the final models for student achievement nested within classrooms and schools were developed, these models were used to assess adjustments to students' predicted achievement that would be suggested when they are taught by a new teacher from a particular university or alternative teacher preparation program. This step was the VAA. This was modeled at the teacher level by a series of codes representing being a new program completer from a particular university. Alternative certification and regular undergraduate programs were modeled separately.

The mean impact of teachers who were recent graduates of particular universities and completers of specific alternative certification 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. The tests for 2006 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 effect for that teacher preparation program in comparison to experienced certified teachers. The number in brackets below the effect is the confidence interval for the effect. The VAA HLM was conducted for both years separately using the same three level model that was developed for each content area for the 2005-2006 year data.

Prior work (Noell, 2006), had used an arbitrary standard of at least 10 new graduates before a program was included in the analyses. However, examination of plots of the variance associated with each program estimate by the number of completers suggests that this is too low a standard for inclusion in analyses. Near that number of graduates, the variance associated with individual programs typically exceeds the variance between program estimates. As a result it would be anticipated that university estimates would be unstable from year to year. However, adopting a higher number of graduates as the threshold creates an alternative problem in that large numbers of programs are excluded from the assessment. For example, in the current data, between 10 and 20 program completers teaching in assessed grades could be identified in mathematics for 9 programs in 2004-2005 and 11 in 2005-2006. Examining the variance versus *n* per program plots for mathematics, science, and social studies for a joint analysis across years (described below), generally, across content areas, the mean ratio of within program variance to between program variance dropped to approximately 0.5 or less at about 25 new graduates. Beyond 50 new graduates per university the ratio generally dropped to approximately 0.2 and stayed at or below that value across content areas.

A fundamental reality of the problem at hand is that for the most part, universities are not going to begin producing dramatically more new teachers than they have in the past. As a result, two different positive outcomes are somewhat irrevocably in conflict. Limiting assessments to programs that have 25 or better yet 50 first and second year teachers in each content area will dramatically limit the programs that can be assessed. This will introduce more selectivity in who is assessed with its attendant undesirable consequences. Including programs that have 12 or 18 graduates will frequently create unstable estimates. At present the most viable solution appears to be pooling data across years to permit the accumulation of observations. The current report is based on doing this across 2004-2005 and 2005-2006. These estimates should, over cohorts, stabilize if teacher preparation contributions and the underlying model parameters are reasonably stable. Analysis of data pooled across years also has the desirable quality of reducing the impact of unusual cohorts of graduates which are much more likely to emerge in smaller preparation programs. Coefficients are only reported herein for TPPs that had at least 25 total graduates based upon the current analyses and at least 10 graduates per year based upon Louisiana's current accountability system for teacher preparation.

Combining Data Across Years

Two general strategies were considered for combining data across years to yield joint estimates. The first strategy employed precision weighted averages. While this should produce reasonable estimates of the coefficients, no obvious generally accepted strategy for arriving at joint standard errors of the coefficients could be identified. Alternatively, the data could be analyzed using combining the two years data sets, but interacting predictors by year so that the model for each year could yield independent coefficients for predictors. The interaction of predictors by year would result in all predictors being set to zero for the year not being estimated (described below). This strategy has the additional advantage that if the codes teacher preparation programs were not interacted with years coefficients and standard errors can be extracted across the two years pooling the data.

In order to combine data across the two years the following analytic structure was used. The dependent variable was the target achievement test score. The predictor variables were the ones that were retained in the analyses described above. In order to allow the models for the two years to remain largely independent, each predictor variable was interacted with year so that all 2004-2005 variables were set to 0 for 2006 test results and all 2005-2006 variables were set to 0 for 2005 test results. In essence the model for each year was allowed to assume coefficients that were as independent of the other year's data to the extent possible. Results were reviewed and the coefficients for variables were quite close to the values that were obtained when each year was analyzed independently.

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 problematic assumption that schools are the same organizational units across years. This is obviously not the case when one considers the impact of redistricting, changes in staff, and the impact of school expansions and contractions. One disadvantage is that the model did not capitalize on the repeated observation of teachers across years. However, no software could be identified 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.

Propensity Score Matching

An additional step was taken to reduce the likelihood that a relatively distinct pattern of new teacher placement would distort the TPP coefficients. Propensity score matching (PSM) was used to select a comparative group of teachers whose class compositions were similar in likelihood to the classes taught by new teachers from each TPP. A particular challenge to PSM in the current context is the use of multilevel models. In the current context, using a 1-to-n strategy for matching nearest neighbors with an *n* of 1, 5, or even 10 will lead to many teachers being the only teacher nested within their school. In these circumstances, estimation of the school level of the model will be poor because most of the information about the teachers in the school will be lost. Models using weighting can have similar consequences. In an HLM context, PSM has the paradoxical risk of introducing bias due to poor estimation of school effects.

To minimize this risk, an approach to PSM was adopted that was designed to maintain as many of the reasonable matches as was possible. The approach used herein was from the procedures described by Rubin and Thomas (2000). Rubin and Thomas describe using the nearest remaining neighbor matching strategy based upon logistic propensity scores using 1 to 1 and 1 to 5 matches. In an alternative they describe using a "coarse" (pp. 574) of caliper .2 to specify a range within which to perform Mahalanobis distance matching. Given the number of programs to match across multiple content areas, the thousands of included teachers, and the desire to maintain a large *n* when propensity matching permitted to maintain a reasonable three level model an adaptation of these two procedures was used. In the current application a fine caliper that was 5% of the width of Rubin and Thomas' coarse caliper was used (i.e., .01). However, all matches within a .01 caliper of any classroom from that specific TPP were retained for analysis.

For each TPP a logistic regression was used to obtain the probability of a classroom being consistent with the classrooms of the new teachers from the TPP. This probability was then converted to a propensity score and all matches within plus or minus a caliper .01 for any classroom taught by a graduate of the TPP was selected for the comparison. Analyses were also run without the PSM for comparison. The correlations for mathematics, science, and social studies between the PSM coefficients and the coefficients for the full data set were $r = .93$, .96, and .97 respectively. For 7 of 44 estimated coefficients the change in coefficient was more than absolute 1 point and in only once case was the change more than 2 points. Generally scores shifted somewhat within the same part of the distribution of scores when PSM was applied. Results presented below are the PSM results.

English Language Arts

Preliminary analyses examining VAA estimates for ELA based upon TPP completion derived from the Department of Education databases suggested a poor relationship between coefficients for universities across years. This was particularly strongly affected by four outlying schools whose coefficients changed dramatically. In contrast, a moderate relationship was evident for mathematics with the correlation for all programs with 10 or more graduates per year of $r = .48$ for the full data and $r = .63$ for the data excluding the hurricane affected districts from both years. The correlation between university coefficients for science ($r = .30$) and social studies ($r = .43$) were somewhat less strong. This degree of consistency is somewhat encouraging given the poor stability that is suggested for individual estimates with fewer than 25 teachers. It may be possible to achieve reasonable stability by accumulating data across multiple years.

More detailed examination of the ELA data clarified that the structure of the student-teacher-course nexus was very different for this content area than for the other content areas. In particular there were strikingly more students with multiple teachers and this was clearly the result of many students having separate reading and English teachers. Exploring the data suggested that analyses separating reading and written language should be pursued prior to reporting ELA results. However, the database that had been constructed to date did not have the specificity of subtests or course designations to permit this separate analysis. Work is currently ongoing to augment the data structure to permit separate analysis of reading and written language in a model that should more closely align with teacher assignment. Results for reading, written language, and/or ELA will not be reported until this work is completed.

Performance Bands for Mathematics, Science, and Social Studies

In order to help readers place data in context, five performance levels were developed in consultation with 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. The performance levels are defined below.

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 a statistically significant difference. The difference between these programs and the mean for new teachers would commonly be statistically significant.

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.

Redesign of Teacher Preparation Programs

One issue that arises in assessing the meaning of the data is the implementation of statewide redesign of teacher preparation in the years immediately prior to the years captured in these analyses. The State's certification structure as well as all existing teacher preparation programs were redesigned in the period of 2000-2003. The years captured in this assessment are 2004-2006. As a result, the bulk of the new teachers captured in this assessment completed TPPs that have since been phased out. The pre-redesign TPPs stopped admitting new students on or before July 1, 2003. As a result the bulk of the data reflect where the programs were immediately prior to implementing their redesigned programs. The pre-redesign data are now out of date and may not reflect the current functioning of the relevant TPPs. These pre-redesign baseline data have been provided to the universities and university systems. These data are provided herein as was requested of the authors.

In three instances, TPPs were identified based upon data provided by the Board of Regents whom the data consisted of 88% to 100% completers of their post-redesign or current programs. In these instances, the results are believed to reflect the functioning of the program that is currently admitting students. These estimates are presented first. It is important to note that these programs are all alternative certification programs. This is an artifact of the fact that alternative certification programs were redesigned first and require less time to complete. The authors and Board of Regents anticipate that the State is at a point where the distribution of program completers should shift decidedly from the pre-redesign to the post-redesign programs. Once programs have a sufficient number of post-redesign program completers, their data will be reported for redesigned programs. It is anticipated that this should be a building and accelerating process over the next one to three years. It is also worth noting that the availability of data for programs prior to redesign may be valuable in examining the effects of redesign.

Tables 15-17 below present the VAA estimates for mathematics, science, and social studies for post-redesign or current programs. The more liberal 68% CI was adopted for this report based on the assumption that for a formative assessment such as this, the consequences of false negatives, failing to identify an exemplary program or one that is struggling, are typically at least comparable to the risks of a false negative.

Table 15: *Teacher Preparation Program Coefficient for* ***Post-Redesign Programs*** *(current): Mathematics*

Level	**Teacher Preparation Program**	**2004-2006 Estimate** (CI)	**Teachers**
1	New Teacher Project ^{NT}	2.1 (0.0, 4.1)	26
2	Northwestern State University Alternative Certification	2.6 (-0.1, 5.3)	49
3	Louisiana College Alternative Certification	-1.6 (-4.8, 1.6)	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 -2.0. Programs noted with the NT superscript were statistically significantly different from new teachers at $p = .05$.

Table 16: *Teacher Preparation Program Coefficient for* ***Post-Redesign Programs*** *(current): Science*

Level	**Teacher Preparation Program**	**2004-2006 Estimate** (CI)	**Teachers**
1	Northwestern State University Alternative Certification [NT]	2.7 (1.2, 4.2)	39
2	Louisiana College Alternative Certification	1.7 (-1.1, 4.6)	25

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 -1.1. Programs noted with the NT superscript were statistically significantly different from new teachers at $p = .05$.

Table 17: *Teacher Preparation Program Coefficient for* ***Post-Redesign Programs*** *(current): Social Studies*

Level	**Teacher Preparation Program**	**2004-2006 Estimate** (CI)	**Teachers**
1	Louisiana College Alternative Certification	5.5 (1.6, 9.4)	28
2	Northwestern State University Alternative Certification	1.6 (-0.4, 3.6)	37

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.

The data for the three alternative certification programs that have graduated new teachers from their redesigned teacher preparation programs is very positive. In six of seven instances the programs fell at level 1 or 2. In two instances the TPP coefficient was statistically significantly distinct from the mean for all new teachers. It is also interesting to note that the coefficient for all new teachers for science was approximately half the magnitude of the coefficient for mathematics or social studies.

Tables 18-20 below present the VAA estimates for mathematics, science, and social studies for pre-redesign or phased out programs.

Table 18: *Teacher Preparation Program Coefficient for* ***Pre-Redesign Programs*** *(phased out programs): Mathematics*

Level	**Teacher Preparation Program**	**2004-2006 Estimate** (CI)	**Teachers**
3	Louisiana State Univ. - Shreveport Undergraduate	-0.1 (-2.5, 2.3)	33
3	University of New Orleans Undergraduate	-0.6 (-2.3, 1.0)	49
3	Nicholls State University Alternative Certification	-0.6 (-2.6, 1.4)	36
3	University of Louisiana – Lafayette Alternative Certification	-0.8 (-2.5, 0.9)	45
3	McNeese State University Undergraduate	-1.3 (-2.6, 0.0)	74
3	University of Louisiana - Lafayette Undergraduate	-1.6 (-2.7, -0.5)	130
3	Louisiana State University Undergraduate	-1.6 (-2.6, -0.6)	124
3	Nicholls State University Undergraduate [ET]	-2.7 (-3.9, -1.5)	53
3	Louisiana Tech University Undergraduate [ET]	-3.4 (-4.8, -1.9)	46
4	Southeastern Louisiana University Undergraduate [ET]	-3.6 (-4.6, -2.5)	115
4	Northwestern State University Undergraduate [ET]	-3.9 (-5.4, -2.3)	42
5	University of Louisiana - Monroe Undergraduate [NT]	-4.9 (-6.3, -3.4)	29

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. Programs noted with the ET superscript were statistically significantly different from experienced teachers at $p = .05$. Programs noted with the NT superscript were statistically significantly different from new teachers at $p = .05$.

Table 19: *Teacher Preparation Program Coefficient for* ***Pre-Redesign Programs*** *(phased out programs): Science*

Level	Teacher Preparation Program	2004-2006 Estimate (CI)	Teachers
3	Louisiana State University Undergraduate	-0.2 (-1.3, 0.9)	98
3	University of Louisiana - Monroe Undergraduate	-0.5 (-2.8, 1.8)	25
3	Louisiana State Univ. - Shreveport Undergraduate	-0.6 (-2.4, 1.2)	35
3	Northwestern State University Undergraduate	-0.8 (-2.0, 0.4)	44
3	Nicholls State University Undergraduate	-1.2 (-2.7, 0.2)	39
3	Southeastern Louisiana University Undergraduate	-1.4 (-2.4, -0.4)	108
3	McNeese State University Undergraduate	-1.5 (-2.8, -0.3)	65
3	University of Louisiana - Lafayette Undergraduate	-1.8 (-2.9, -0.7)	117
4	University of Louisiana - Lafayette Alternative Certification	-2.6 (-4.0, -1.2)	34
4	University of New Orleans Undergraduate	-3.0 (-4.7, -1.3)	41
5	Louisiana Tech University Undergraduate ^{NT}	-6.4 (-8.4, -4.4)	33

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 -1.1. Programs noted with the NT superscript were statistically significantly different from new teachers at $p = .05$.

Table 20: *Teacher Preparation Program Coefficient for* ***Pre-Redesign Programs*** *(phased out programs): Social Studies*

Level	**Teacher Preparation Program**	**2004-2006 Estimate** (CI)	**Teachers**
2	Louisiana State University Undergraduate NT	-0.1 (-1.1, 0.8)	111
3	University of New Orleans Undergraduate	-0.4 (-2.3, 1.5)	45
3	Northwestern State University Undergraduate	-1.0 (-3.0, 0.9)	45
3	Nicholls State University Undergraduate	-1.6 (-3.2, 0.0)	46
3	Nicholls State University Alternative Certification	-2.2 (-4.2, -0.3)	31
3	Southeastern Louisiana University Undergraduate	-2.3 (-3.5, -1.2)	102
3	McNeese State University Undergraduate	-2.4 (-3.7, -1.2)	68
3	Louisiana State Univ. - Shreveport Undergraduate	-3.4 (-5.9, -1.0)	40
3	University of Louisiana - Monroe Undergraduate	-3.9 (-6.1, -1.7)	27
3	Southern University Undergraduate	-4.3 (-6.7, -1.9)	35
4	University of Louisiana - Lafayette Undergraduate ET	-4.2 (-5.3, -3.0)	127
4	University of Louisiana - Lafayette Alternative Certification	-4.8 (-7.3, -2.3)	36
4	Louisiana Tech University Undergraduate ET	-5.6 (-7.9, -3.2)	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 -2.1. Programs noted with the ET superscript were statistically significantly different from experienced teachers at p = .05. Programs noted with the NT superscript were statistically significantly different from new teachers at p = .05.

The data for the pre-redesign programs present a different picture than those for the post-redesign programs. First, nearly all of these programs are undergraduate programs. That is not surprising with the realities that they were generally implemented after the alternative certification programs and take longer to complete. Second, no level 1 and only one level 2 program was identified. Third, 25% of the programs fell within levels 4 or 5, with two programs falling statistically significantly below the mean for all new teachers in that content area. It is also interesting to note that both consistency and inconsistency was evident within universities across content areas. For example, LSU-Baton Rouge had the only level 2 program in social studies education and was quite close to that same break point in science. In contrast, in mathematics the TPP coefficient for LSU was close to the coefficient for all new teachers.

Summary

Examining the results across content areas suggests some preliminary conclusions. First, with as few as 10 teachers per observed cohort, somewhat encouraging results for the consistency of university estimates were obtained for mathematics, science, and social studies. Given the possibility of strengthening those estimates by accumulating successive cohort observations and achieving a reasonable ratio of within to between variance, it appears that it is possible to obtain estimates that will achieve reasonable stability through multiple years of pooled data. Second, although it should come as no surprise that most programs clustered around the estimated coefficient for all new teachers, sufficient variability in TPP coefficients was evident to identify a few programs as outlying to varying degrees in comparison to either new teachers or experienced teachers. In a minority of cases these differences were statistically significant.

A third and striking finding was the degree to which redesigned programs outperformed pre-redesign programs. However, so many differences exist between the post-redesign and pre-redesign programs (e.g., predominantly undergraduate versus alternative certification programs) that it is difficult to interpret this finding. It is interesting to note that the alternative certification programs that have not yet begun to produce large numbers of post-redesign graduates did not outperform the undergraduate programs. Over the next one to two years as substantial numbers of graduates of redesigned undergraduate teacher preparation programs enter the work force it will be possible to reexamine this issue. Although the data suggest that enough variability exists in TPP coefficients that it may be possible to identify a minority of programs as doing particularly well or poorly, it is important to reiterate that the bulk of the coefficients available today reflect programs prior to redesign and may not reflect the programs that currently admit students.

VIII. Additional Detailed Family Demographic Variables

One concern that has arisen regarding the use of student achievement data to assess TPP efficacy is that it will not consider one of the major determinants of student achievement that is beyond the control of schools: families. This is a reasoned concern that reflects the wealth of data demonstrating that student achievement is correlated with a variety of family demographic factors, parenting practices, and family resources (Coleman, 1989; Downer & Pianta, 2006; Hill & Craft, 2003; White, 1982). An alternative argument has been advanced that although family factors are powerful influences on educational attainment, their influence is already expressed to a considerable degree in the existing educational databases (Ballou et al., 2004).

It may be that neither argument is generally correct, but that they represent outcomes at different ends of the distribution of educational data systems. Specifically, it may be the case that when educational data are sparse, assessments are weakly related to one another, or that supplementary data such as attendance and special education disability status are not available that family data would substantially improve the prediction of student achievement. In contrast, it is also the case that the variability in student achievement is finite and that a great many predictors share overlapping variance. If this is the case, and it certainly appears to be, then a reasonable subset of those predictors may achieve adequate prediction while omitting other variables that are developmentally important, but share variance with variables that are already being employed as predictors.

The critical consideration for this research effort and Louisiana's efforts to assess TPP efficacy is not the degree to which family variables would improve prediction in the abstract, but the degree to which they would improve on the prediction of achievement given the specific tests and demographic variables that are available. Stated differently, to what extent do the prior tests with their specific technical characteristics and student demographic variables share the same variance in predicting student achievement as family variables such as parental educational attainment, marital status, or parental engagement with their children's education? If the degree is high, it would suggest that family influences, although not measured, are already evident as effects in prior achievement.

One critical challenge in addressing this question is that it is a substantial missing data problem. The types of data that are of interest are not commonly available to educational researchers. As part of the VAA research project at LSU a data collection effort was undertaken to examine the extent to which additional data from families would improve the prediction of student achievement. In the interest of space the following briefly summarizes the research methods and key findings that are related to the VAA study.

Participants. A stratified random sample of schools was identified and recruited to participate in the family survey data collection. Initially all schools that included grades 4-9 were eligible to participate. However, subsequent to the landfalls of Hurricanes Rita and Katrina, the sample was revised and the school districts of landfall and immediately adjoining school districts for Katrina were excluded. Additionally, when principals were contacted to solicit their participation, they were asked if more than 10% of their students had been displaced by the hurricanes. Any school in which more than 10% of the student body was hurricane displaced was dropped and replaced.

The sample was stratified such that nine schools were recruited to represent each of three segments of the demographic distribution of schools in Louisiana. The segments were low, at or below the 25th percentile on the variables of interest; middle, the 26th - 75th percentile, and high, the 76th – 99th percentile. Nine schools were randomly selected that represented the low, middle, and high end of the distribution of each variable. This sampling plan was intentionally mildly skewed to the tails of the distribution to increase the probability that the sample included sufficient representation of the range of variability evident in Louisiana, rather than being purely representative.

The variables that were used to stratify the sample were the percentage of students who were ethnic minorities, percentage of students receiving free or reduced price lunch, and the school performance score (SPS). The SPS is Louisiana's school accountability index that is comprised of weights of several variables, but the weighting is strongly dominated by achievement test data. Schools with high SPS will have students who are performing relatively well on the state assessments. A final additional demographic variable that was used for stratification was the locale code from the U.S. census for the school's zip code. Three strata were identified: urban (mid-size city code), suburban (two different urban fringe codes), and rural. Schools were offered $1500 to participate in data collection. Several schools that were contacted declined and replacement schools were identified through random selection.

Method. Schools that agreed to participate distributed survey packets to families by sending them home with students. The packet contained a cover letter soliciting participation and providing informed consent information, a brief survey (one page), and a sealable envelope in which to return the survey to the school. Parents were asked to complete the survey, place it in the envelope, seal the envelope, and return it to school with their son or daughter. Once at school, sealed envelopes were placed in a central storage container until they were retrieved by the research team.

The survey asked the student's first name, last name, date of birth, gender, grade level, and ethnicity. These data, along with the school the student was enrolled in were used to link survey data to achievement records. A series of questions were derived from prior reviews of educational research identifying family variables that have been identified as predicting student achievement. The survey questions and response options are presented in Table 19.

Table 19: *Family Demographic Survey Items*

Item	Response Options
Has the child always lived with the same parent (mom OR dad) since they were born?	Yes No
Has the child always lived with the same 2 parents (mom AND dad) since they were born?	Yes No
Marital status	Married Separated Divorced Never married Widowed
Number of adults that live in the home	Grid Number
Number of children that live in the home	Grid Number
Age of mother when child was born	Grid Number
Education level: mother	8^{th} grade or less Some high school High school diploma Some college (at least 1 year) Vocational technical training College graduate Graduate professional degree
Education level: father	Same as mother
Annual family income	0-4,999 5,000-9,999 10,000-19,999 20,000-39,999 40,000 & up[1]
Number of times your family has moved since your child has been in school	0 1 2 3 or more
Number of times your child has changed schools since kindergarten	0 1 2 3 or more
How many activities outside of the home is your child involved in?	0 1 2 3 or more
How many of your child's friends' parents do you know?	0 1 2 3 or more
Do you have a working computer in your home?	Yes No
How many minutes per week do you spend with your child helping or talking about school work?	0-15 min 15-30 min 30 min-1 hour 1 hour +
How many times in a year do you visit your child's school for an event (do NOT include teacher conferences).	0 1 2 3 or more

Table note. 1. Due to a communication error the maximum value for family income was lower than would be desirable.

Response Rate. Despite follow-up contacts and site visits by the research team only 18 of the 27 schools that agreed to collect data returned surveys. Typically principals cited the busy end of year schedule, including standardized testing requirements, as having interfered with planned data collection. Chi-square analyses for responding schools versus schools not returning data revealed no significant differences on any of the four stratification variables. Although all strata were represented for each demographic variable, schools that were low in poverty, high in prior achievement, and low in minority enrollment were somewhat disproportionately represented. They ranged from 37.5 – 43.8% rather than the targeted 33.3%.

Survey responses were provided by the parents of 1512 students and of these responses 1283 (85%) were matched to student achievement records. These matched responses included 23.5% of the students in the assessed grades in the responding schools. Students whose parent returned a survey were approximately equally distributed in 4^{th} through 8^{th} grade. Only 2% of the students whose parents responded were in 9^{th} grade. The students whose parents responded were demographically distinct from the population of students contributing to the data analyses for Louisiana despite the sampling plan. They were more often Caucasian (73.3% versus 50.4%), less often in special education (6.9% versus 13.1%), more often gifted (11.1% versus 4.6%), less often receiving free/reduced price lunch (40.7% versus 64.2%), and more often female (56% versus 48.2%). As a whole the students whose parents responded can be described as more Caucasian, more educationally advantaged, more economically advantaged, and more female than the State.

Results. The central research question for this data collection was the extent to which additional information from families would contribute variance in predicting student achievement that was unique from that contained in the educational databases available. As an initial step the regression models for all available prior achievement data for one year and demographic variables described in Section III were initially repeated with the sample of responders to examine the degree to which their data matched prior results for shared variance.

Analyses revealed that the educational databases yielded more precise predictions for the sample of responders than it did for the State. For ELA, the adjusted shared variance was notably higher at .71 for the sample versus .66 for the State. In mathematics the adjusted shared variance was .68 while for the State the result had been .65. These results demonstrate that the selectivity in responding that is evident in the demographics created a sample that is mildly more orderly in terms of the ability of school based variables' to predict achievement.

The next stage of the analysis examined the ability of the family demographic survey to improve prediction of student achievement. Variables coded as yes or no were coded 1 for yes and 0 for no. Continuous variables (e.g., mother's age at student's birth) were entered as they were coded. Ordinal variables were coded from 1 to the top of their respective scale. Marital status was dichotomized into married and not married. The number of adults and children in the home was converted to the ratio of adults to children living in the home. Although the overall rate of missing data was low for survey responses (3% of total survey response), 31% of cases contained at least one missing data point for the survey. In order to make use of the maximum amount of information in the data set, multiple imputation for the missing data was used to create three imputations. For the imputation process, all of the survey data, the achievement data, and the school demographic variables were employed.

Results across imputations were strikingly similar and the results presented below are the mean results across the 3 imputations. The regressions described above were repeated for the imputed dataset and following entry of the school-based variables, the family demographic variables were then forced in as a block. The adjusted shared variance did not increase with the addition of the family variables to a degree that is likely to have much educational significance (.710 to .715). The increase in variance accounted for in mathematics was even less (.682 to .684). Although some exploratory work was done examining the family predictors and the strength of their relationship to student achievement that is not central to the work at hand will not be reported here in the interest of space. The authors also note that this is exactly the sort of problem for which a stepwise method (variable selection issue) is intuitively appealing, but given the well documented capitalization on chance, the ambiguity of the concept of best set of variables, and the very small amount of variance that family variables appear to explain beyond that shared with achievement, it is doubtful that this intuitively appealing method would produce replicable results (Thompson, 1995).

Summary. Within the limits of the family survey employed and the selectivity of responding, these data suggest that Louisiana's extensive education database accounts for the bulk of the variance shared between family variables and student achievement. If these data hold for the larger population, the additional variance added by including family variables was sufficiently modest that it would be exceedingly unlikely to change the VAA of TPP in a substantive way.

The research team will have an opportunity to replicate this aspect of the work for the 2006-2007 school year. The survey was repeated and based on experiences with the data collection for 2005-2006 modifications were made to the coordination with schools that increased the response rate.

IX. Reliability of the Curriculum Database

An additional concern that has been raised within Louisiana is that the curriculum database that links teachers, students, and courses may not be sufficiently accurate to support the VAA of TPP work. In order to assess this issue the 27 schools that had been sampled to participate in the family survey (described above) were asked to complete a curriculum survey at the end of the school year. Twenty of these schools agreed to do so and returned the data. The surveys presented the curriculum reports that the schools had provided in the fall and asked that whether the teacher identified as teaching each class was still at the school, whether the teacher taught the class indicated in the curriculum record, and whether the teacher had left the school. For each student, schools reported the following: if the student was still being taught by that teacher in that subject, if the student had changed teachers within the school, and if the student had left the school.

Curriculum surveys were returned reflecting 19,626 student teacher linkages in tested subjects and grades. Students typically provided multiple linkages because of enrollment in multiple classes. Returned surveys indicated that in 5.2% of cases the teacher did not teach that subject. This error applied to 4.6% of students. In 4.3% of cases, schools reported that the teacher of record had left the school. As a result of these data, additional data were obtained from the Louisiana Department of Education from which it was possible to identify teachers who changed schools or left teaching during the year. For all VAA reported herein, those teachers were dropped. Of the remaining students whose teachers were still teaching the target class at the school (90.6% of teachers), 2.4% of students had changed teachers, but were still in the school and 7.4% had left the school.

The end of year curriculum survey identified several types of errors that can occur in assuming that the teacher of record in the curriculum database is the teacher who taught that student and obtained the rate at which these errors occurred in this sample. The first type of error occurs when students' class enrollment status changed after the database was completed. In the majority of cases this occurred when students or teachers left the school. Screening for both of these events are now routinely included in the VAA analyses. Although dropping these records increases lost data, it reduces misattribution of students to teachers. The other change induced error occurred when students changed teachers within the school. This occurred for 2.4% of the students whose teachers remained at the school in the reported class the entire year. Although it would be desirable to eliminate this error, it appears to be of relatively small magnitude. Data that can address this issue are not currently available.

The remaining error is the most troublesome. In 5% of cases schools reported that the teacher was not teaching the course that they had been identified as teaching at the beginning of the year. Unlike the mobility issues, there is no way to detect this error without a second curriculum data collection at the end of the year and/or assuring greater care in the original entry of data. This type of error will introduce additional error variance into the VAA estimates and contributes to the need for accumulating more teacher observations prior to estimating TPP coefficients.

X. Teacher ACT Scores at College Admission and Effectiveness

The research plan included examining the relationship between teachers' ACT scores prior to entering university education and their effectiveness. At present work is ongoing at the Board of Regents to assemble a more complete ACT database that will permit this analysis that may shed some light on the relative impact on admissions on VAA estimates for TPPs.

XI. Summary

Analyses were conducted to replicate and extend the prior statewide analyses for the 2004-2005 school year (Noell, 2006). Construction of the longitudinal database suggested that a sufficient quantity and quality of data appear to be available to support longitudinal analysis of educational inputs such as teacher preparation. For example, the 93% linkage rate for student data across years was very encouraging. However, it is important to note that a number of practical realities resulted in the loss of a considerable amount of data despite a high rate of matching students across years and teachers to students. The most critical issue was mobility with students entering public school for only part of the year, changing schools during the year, teachers leaving teaching, and teachers changing schools. Additionally, students who were repeating the same grade were dropped because the meaning of their assessment data is different from students who were promoted. In total, mobility, the failure to match, and retention resulted in the exclusion of 37% of students' records from the analyses. As a result, what these analyses represent is the efficacy of teachers who remain in one school for the year teaching 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.

Following the construction of the database, a number of analyses were conducted. First, preliminary statewide OLS regression analyses were conducted examining the use of various predictors of achievement. Next, mixed linear models of student achievement were developed for each content area using student level, classroom level, and school building variables to predict achievement. These models nested students within classrooms and classrooms within schools. Third, these models were then applied in each content area to implement a VAA of teacher preparation. The same models were fit to 2004-2005 and 2005-2006 school year data. Fourth, OLS regression analyses were conducted examining the increment in variance in student achievement that could be predicted by incorporating additional family variables from a sample of families responding to a demographic survey. Finally, an end of year survey was conducted to examine the accuracy of the curriculum survey data used to link students and teachers in 20 schools.

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

1. The ordinary least squares regressions demonstrate a strong relationship between prior year achievement and current year achievement in the content area. Adding achievement in the three other domains strengthened that relationship as did adding student level demographics and attendance data.

2. The data suggest that the hurricane disrupted codes available in the educational databases had little statistical power in predicting student achievement. With the general decrease in available records and the pattern of data related to the hurricane disrupted codes, to a large degree it appears that the students whose education was substantially disrupted by Hurricanes Katrina and Rita are not represented in the database. This is particularly true when students and teachers who change schools once the school year has started are dropped from analyses, as is the case in the VAA analyses reported herein.

3. Examination of partitioning of variance once again supported a nesting structure with students nested within classrooms nested within schools.

4. 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. Variables entering at the classroom and school level were somewhat consistent,

but on the whole accounted for a relatively small portion of total variance in student achievement.

5. Examination of teacher effects by years experience across all four content areas and both academic years produced somewhat mixed results. The data typically suggested that teachers in either their first 2 or 3 years of teaching might be described as new. Based on the available data the decision was made to maintain the working operational definition of new teachers as first and second year teachers.

6. VAA of TPP was conducted across both the 2004-2005 and 2005-2006 academic years. Examination of the variance within university estimates relative to variance between estimates suggested that the threshold of at least 25 teachers would be needed to obtain reasonably stable estimates.

7. Data using a lower bound of 10 graduates per year suggested some stability in TPP coefficients across years in science, social studies, and mathematics with correlations ranging from .3 to .48. Unfortunately, setting a boundary of 25 graduates per year, which is the more reasonable standard, results in only a few programs contributing to the analysis. However, the ability to pool data across years should increase the number of graduates per TPP and achieve reasonably stable estimates.

8. The results suggest that it is possible to identify some TPPs within each content area that are outlying to varying degrees using either new teachers or experienced teachers as an anchor. Results were more positive for the minority of programs that had sufficient data regarding post-redesign program completers. However, results were available for three only alternative certification programs. Results for the pre-redesign programs were more mixed. Although these data reflect programs that have been phased out, they will serve as a useful basis for comparison for examining the impact of redesign.

9. Examination of family demographic data in a sample of 1283 students found that the demographic variables increased variance shared with student achievement test scores to a very small degree that was unlikely to substantively affect the VAA of TPP. This is not to argue that family factors are unimportant; clearly, they are. They may simply share so much variance with the data already in the educational databases that they would add little to the assessment. The authors do not presume that this is a resolved issue, only that the available evidence suggests that additional family data would not substantially strengthen the assessment of TPPs. This is an issue that will be re-examined for 2006-2007 with a larger sample of students.

10. The end of year curriculum survey found schools confirmed the vast majority of student – teacher – curriculum links. Among those that were reported to be

inaccurate at the end of the year, the majority of those were due to either student or teacher mobility and could be detected with other databases (attendance and employment). The remaining errors occurred because teachers were reported as not teaching that subject, despite still being in the school, or student mobility between teachers within the school. The current data collection of student – teacher – curriculum links and associated databases cannot detect or correct these errors. They occur at a fairly modest level, but could introduce error for a smaller TPP whose graduate cohorts are small. The only practical way to address this would be a second curriculum database collection at the end of the school year. Given the administrative demands that would create and the fact that the State does not currently have another need to collect that data a second time, it appears that a second data collection is unlikely at present.

In summary, the data suggest that with data pooled across years and a reasonable threshold for the number of observations of teachers who must be present in order to report findings, it should be possible to produce VAA estimates for TPP that are reasonably stable. Additionally, the data suggest that in some content areas there are TPPs that will be outliers as producing improved student outcomes or poorer student outcomes. The truly interesting work lies just ahead. What use can be made of the data to strengthen teacher preparation in Louisiana so that new teachers entering the labor market are more effective than the teachers that came before them? Will the data suggest that redesign has already strengthened the preparation of new teachers at some universities? If these questions can be answered, they will contribute to the larger policy goal within Louisiana of improving the quality of education provided to all of Louisiana's sons and daughters.

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.

Coleman, J. S. (1989). The family, the community, and the future of education. In W. J. Weston (Ed.), *Education and the American family: A research synthesis.* (pp. 169-185). New York University Press: New York.

Downer, J. T., & Pianta, R. C. (2006). Academic and cognitive functioning in first grade: Associations with earlier home and child care predictors and with concurrent home and classroom experiences. *School Psychology Review, 35,* 11-30.

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.

Hill, N. E., & Craft, S. A. (2003). Parent-school involvement and school performance: Mediated pathways among socioeconomically comparable African American and Euro-American families. *Journal of Educational Psychology, 95,* 74-83.

Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual level heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics, 1,* 223-252.

Martineau, J. A. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value added accountability. *Journal of Educational and Behavioral Statistics, 31,* 35-62.

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. (2004). *Technical report of: Assessing Teacher preparation program effectiveness: A pilot examination of value added approaches.* Retrieved from http://asa.regents.state.la.us/TE/technical_report.pdf

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. & Burns, J. L. (2006). Value added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education, 57,* 37-50.

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.

Roderick, M., Jaccob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation & Policy Analysis, 24,* 333-357.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the *Prospects* study of elementary schools. *Teachers College Record, 104,* 1525-1567.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95,* 573-585.

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,* F3-F33.

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.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55,* 525-534.

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

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91,* 461-481.