The Status of the Development of the Value Added Assessment Model as specified in Act 54

A report to the Senate Education Committee and the House Education Committee of the Louisiana Legislature

February 25, 2011

Table of Contents

Executive Summary

Four developmental processes were deployed in support of the implementation of the value added model required under Act 54. A statewide advisory panel was formed that includes diverse representation from across the State including legislators with the majority of the members being practicing teachers. This panel's review and advising role is ongoing. The second major process was the development, testing, and deployment of a secure web portal through which teachers and educational leaders are able to verify the accuracy of class rosters before they contribute to value added analysis and through which they can access the results. The third major process was the field testing of the process for providing value added results to teachers. This occurred in 19 volunteer districts to which professional development was provided to teachers and leaders. Educators in these districts were provided with professional development and materials to prepare them to interpret their scores. They were also provided with access to their scores for 2009-2010. Follow-up activities with these districts are underway.

The fourth major developmental activity has been the analytic work to prepare the results that are shared with the teachers. This work has examined the impact of a number of model design choices that are, have been, or will be reviewed by the State advisory panel. This report provides detailed information regarding the calculation method and highlights key findings. The authors have interpreted the data presented here, combined with additional data to suggest the inclusion of some factors beyond prior achievement. Disability diagnosis is advised, as is the inclusion of classroom composition variables.

Notable among the findings is the result that there is a group of teachers who were consistently in either the lowest performing or the highest performing group of teachers across years. Consistent cross year results, when they are evident for a teacher, appear to provide a basis for engaging in substantive work to improve outcomes for the students of the lowest performing teachers and efforts to retain the highest performing teachers. An encouraging finding is that cross year consistency is improving as the data quality is enhanced.

Processes Supporting Development of the Value Added Model

Four processes were deployed in support of the development of the value added model. First, pursuant to Act 54, the Superintendent of Education convened the Advisory Committee for Educator Evaluation (ACEE). That group has met and continues to meet on an ongoing basis to receive information about the provisions of Act 54, potential implementation strategies, the implications of those strategies, and develop recommendations to BESE regarding the implementation of Act 54. ACEE has met twice, with upcoming meetings scheduled for February and March 2011. This review and advisory committee includes diverse representation from across the State including legislators with the majority of the committee is made up of practicing teachers.

Second, the Louisiana Department of Education has developed and deployed the Curriculum Verification and Reporting Portal (CVR). The CVR provides a secure online site where teachers can verify the accuracy of their student rosters and class schedules before these data are used to contribute to their value added assessment. The CVR was developed to address two key concerns. The first key concern is that observation by a number of scholars that data quality has remained a critical barrier to accurately estimating teacher contributions to student progress and the consistency of that contribution. The second key concern is the need to create as much transparency as possible into the process for deriving value added scores. With the deployment of the CVR, teachers have the opportunity to know exactly which students are contributing to their results and correct data errors. The CVR also allows teachers, principals, and district superintendents can access the value added results. Generally, the CVR portal is simple enough and follows common web convention to the extent that it would be expected that most teachers would be able to use the portal without formal instruction. Live online training is provided for using the CVR's features for educators who would like it. Technical support is provided for both data review and during the statewide roster verification period.

The third process supporting the value added component of Act 54 has been the field testing of the educator professional development materials, CVR, and results with 19 volunteer school districts and two charter schools. This professional development included meeting with district superintendents, principals, and teacher leaders from participating schools and districts. During the professional development educators were provided a briefing on value added in a small group format that included the opportunity for discussion and questions. They were provided with training materials for redelivery of the session in their home schools including a PowerPoint® presentation, a video, and printed materials. In addition they were provided with follow up resources for questions that arose that they could not answer. Depending on the size of the district, from 1 to 24 professional development sessions were held.

The participating schools' value added results were uploaded approximately 2 to 3 weeks following the initial training to permit remaining teachers to receive the information prior to having their scores. Follow-up meetings have been held with a number of schools and districts to discuss results, concerns, and data. The LDOE team will conduct additional focus groups with an additional portion of the participating schools. The table below provides the district names and the number of schools within that district that participated in the field test.

Table 1. Districts Participating in the Field Test

School District/Organization	Schools
Ascension	27
Baker	3
DeSoto	10
East Baton Rouge	10
East Feliciana	8
Iberville	8
Jefferson	89
Lafourche	24
Monroe City	22
Recovery	22
Richland	10
Sabine	13
St. Helena	2
St. James	9
St. John	12
St. Martin	13
Terrebonne	33
West Baton Rouge	7
West Feliciana	4
La Assoc. of Charter Schools	2
Total	328

The fourth process supporting deployment of the value added assessment is the analytic work that has been used to derive the results provided to the teachers. The analytic work was conducted by LDOE staff led by two PhD level researchers with extensive experience with value added models and their application to data in Louisiana. The balance of this document describes the analytic process and some of its key outcomes.

I. Technical Process and Findings

1. Introduction

This technical brief summarizes the pilot examination of student-teacher achievement outcomes for the 2009-2010 school year that were shared with teachers in 328 field test schools during the 2010-2011 school year. Outcomes were assessed via a value added model. The assessment used regression of student data (achievement, demographics, and attendance) to estimate typical student achievement for students with the same background characteristics and then compare typical outcomes to actual outcomes.

In the context of this report, *value added analysis* (VAA) describes the use of demographic, discipline, attendance, and prior achievement history to estimate typical outcomes for students in a specific content domain (e.g., Mathematics) based on a longitudinal data set derived from all students who took state mandated tests in grades 3 through 9 in Louisiana. The assessment uses a relatively complex model that includes the grouping of students within classrooms.

The current model, where feasible, was developed to address concerns raised by researchers and policy makers regarding variable selection/inclusion and data quality as they emerge in the application of value added models. This included the use of a model process that permitted the inclusion of all students with prior achievement data (described below). Due to low levels of test non-participation in Louisiana this results in a substantially more complete database than is commonly available. The predictor variables were expanded to include non-test variables such as attendance, disability diagnosis, and discipline history. The predictor variables were also expanded to include class composition variables to attend to peer influences on achievement. The CVR was deployed to assure the accuracy of teacher rosters; generally, the data quality in Louisiana has the advantage of having been continuously improved over the last decade due to high-stakes accountability.

2. Database Merging Process

Data were drawn from the standardized test files (*i*LEAP and *LEAP-21*) for spring 2007, 2008, 2009, and 2010; the Louisiana Educational Accountability Data System (LEADS) linking students to teachers; and supplemental student databases. Data analyses for 2007-2008 and 2008-2009 were also conducted to supplement the current year work and provide a point of comparison. The testing and supplemental databases provided data regarding attendance, enrollment, disability diagnosis, limited English proficiency, free lunch status, reduced price lunch, Section 504 status, disciplinary infractions, and demographic variables (e.g., race and gender). Data regarding teachers were drawn from the certification database, teacher attendance, and teacher demographic databases. A multistage process was used to create longitudinal records for students describing achievement, attendance, and demographic factors across years. The student and teacher databases were then linked through LEADS.

Initially, duplicate records and multiple partially complete records that described the same student within separate databases were resolved. 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 the student identification number system that was developed previously by the Strategic Research and Analysis (SRAA) unit at the Louisiana Department of Education. Details of this process are available from SRAA. Table 2 presents the number of records available in each content area.

Table 2. Students and Teachers Available Overall and in Each Content Area

	Overall	**English-Language Arts**	**Reading**	**Mathematics**	**Science**	**Social Studies**
Students	257,252	249,588	173,816	249,382	210,429	207,638
Teachers	15,691	7,939	6,216	7,013	5,299	5,724

Several important decision points are noteworthy. Initial records were limited to students who completed one assessment in grades 4-9 to permit the availability of one year prior achievement data. The testing program begins in the 3rd grade, so 4th graders would have their matched 3rd grade achievement data as predictors of 4th grade achievement. In order to be included in the analyses, a student was required to be enrolled in the same school from September 15, 2008 to March 15, 2009. These dates were set by the field test team. Prior to Act 54 reaching full implementation, the Board of Elementary and Secondary Education (BESE) will have to set the required dates of enrollment for a student to be included. Because the student-teacher-course nexus data are collected only once per year, once a student changes schools within that time period, it is not possible to ascribe achievement measured at the end of that period to a particular teacher. The records available for analysis were attenuated for reading by the reality that few students have an identifiable reading teacher after the 6th grade. The students available for assessment in science and social studies were attenuated because the 9th grade assessment does not include these subjects. Finally, in order to be included in the analyses, the students' attendance and achievement records had to be matched to the LEADS curriculum data to identify which courses the students took and who taught those courses. Additionally, the attendance and course databases were used to confirm that the student was enrolled in the same site.

Course codes were collapsed into groups that were associated with specific test areas (ELA, reading, mathematics, science, social studies). Courses that do not fit these specific test areas, such as band, are dropped from the database.

It is important to note that the first full statewide deployment of the CVR occurred in spring 2010. The comparative analyses between years described below are based on unverified rosters for 2007-2008 and 2008-2009. It is the authors' hypothesis that when two years of verified rosters are available, the relationship between consecutive years may be strengthened as error variance associated with inaccurate student-teacher links is removed.

Additional work was conducted to complete the datasets. Student achievement scores were re-standardized to mean of 300 and standard deviation of 50 across grade and promotional paths. These values were selected because they closely approximate the typical mean and standard deviation of Louisiana's assessments across grades and years. When re-standardizing, the content scaled score was used. Promotional paths refer to how many consecutive years a student had been promoted and have predictor data (i.e., Path 3 means the student was promoted 3 consecutive years; Path 2 means the student was promoted 2 consecutive years, and so on). See Figure 1 for a graphical display of promotional paths. Table 3 describes the number of students in each path for each content area. This process of standardization using paths was adopted for three reasons. First, it allowed retention of all student records with at least two consecutive years of testing. Second, the approach takes students' promotion histories into account. Third, it addressed a phenomenon that emerged in the data in which teachers in specific grade levels appeared to be systematically more or less effective than teachers in neighboring grades and the phenomenon appeared to be attributable to the pattern of promotions and retention being grade specific. For example, there is a higher rate of retention in 4^{th} grade than any other grade level in the assessed span due to high stakes testing in 4^{th} grade. Additionally, re-standardization was also required by the social context of test administration. For example, 8^{th} grade is a high-stakes examination year in which promotion to high school is dependent on test performance. There is a consistent (across students and years) positive shift in performance in the 8^{th} grade compared to all neighboring grades. Failure to attend to this phenomenon would result in teachers in the 7^{th} and 9^{th} grades being consistently found to be substantially less effective than teachers in the 8^{th} grade as a result of the social consequences of the test.

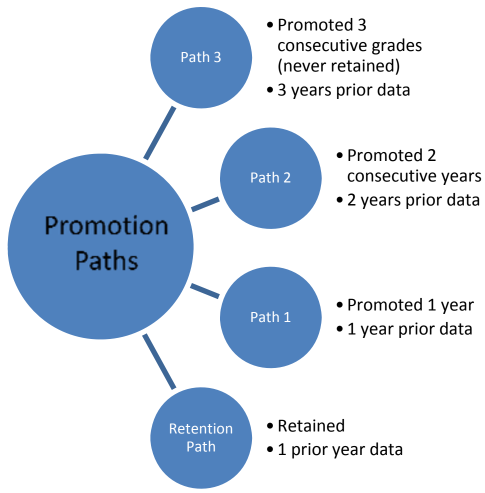


Figure 1. Diagram of promotional paths

Table 3. Number of Students in Each Promotional Path by Content Area

	English-Language Arts	Reading	Mathematics	Science	Social Studies
Path 3	125,967	72,247	125,918	97,392	96,460
Path 2	47,980	40,544	48,045	45,679	45,472
Path 1	63,436	55,703	63,276	59,604	59,300
Retention Path	12,205	9,106	12,143	10,431	10,343

Indicator variables were created to identify student characteristics as well. Indicator codes identify student characteristics using 0s and 1s. If a student has a 1 for an indicator variable it means the student has this characteristic. Indicator codes were used to identify students who were identified as members of the following special education disability groups: emotionally disturbed, specific learning disabled, mildly mentally disabled, speech/language disabled, other health impaired, or other special education disability. Additionally, indicator codes were used for limited English proficiency, Section 504 status, gender, receive free lunch, receive reduced lunch, and ethnicity classification (each ethnic category received its own indicator code).

The final data structure contained a number of variables used to estimate typical student achievement outcomes and links students to teachers based on the course. Table 4 displays the variables used in analyses that were included in the databases.

Table 4. Student Level Variables Retained in the Field Test Model (pre ACEE recommendation and BESE policy)

Variable
Emotionally Disturbed
Speech and Language Disability
Mild Mental Retardation
Specific Learning Disability
Other Health Impaired
Special Education - Other
Gifted
Section 504
Free Lunch
Reduced Price Lunch
Limited English Proficiency
Student Absences
Suspensions (prior year)
Expulsions (prior year)
Prior Mathematics Test (1-3 years based on path)
Prior Reading Test (1-3 years based on path)
Prior Science Test (1-3 years based on path)
Prior Social Studies Test (1-3 years based on path)
Prior English-Language Arts Test (1-3 years based on path)
Squares and Cubes of All Prior Achievement Predictors

3. Value Added Analysis

Once the databases were constructed, the assessment of student-teacher achievement outcomes was calculated as follows. Students who had multiple teachers in a content area were retained in the dataset for their promotional path for each teacher, but were weighted in proportion to the number of teachers they had in that subject. So for example, if a student had two mathematics teachers, the student would have a 0.5 weight in contributing to each teacher's assessment result. Analyses for each content area were conducted separately. The analysis was conducted in three steps. The first two steps were implemented separately for each promotion path and the final step brought all of the data together to obtain student-teacher achievement outcomes.

Step 1. In the first step, data within each path were analyzed using a regression model with classroom centering to obtain the regression coefficients for each predictor. One of the challenges associated with deriving predictor coefficients is accounting for the possibility that the predictors are correlated with teacher efficacy. For example, it is possible that economically disadvantaged students systematically receive less well prepared or less effective teachers. In order to provide a statistical control for this possibility, this stage of the analysis was conducted with classroom centering to obtain the coefficients. This is functionally equivalent to entering teacher fixed effects. As a result the coefficients that were obtained for the predictors would be uncorrelated with (be orthogonal to) teacher effects. Separate intercepts were derived for each grade level.

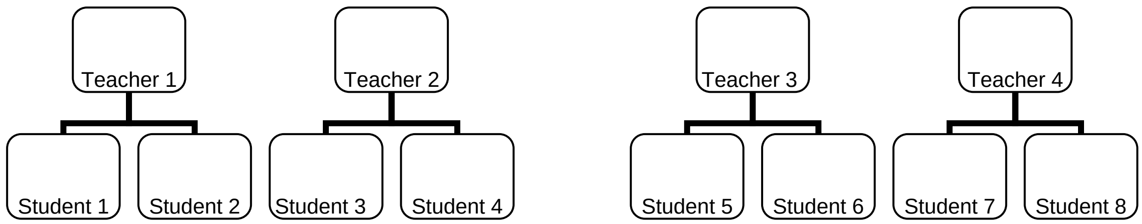
The possibility of crossing grade by path to obtain unique path by path coefficients was examined and did not appear to be viable due to the small number of students with some of the low incidence predictors in some of the very low population paths. In some atypical paths (e.g., 7^{th} grade students with only one year of predictor data) there might be only 0, 1, or 2 students with a specific disability opening up the possibility to severely distorted and unstable coefficients.

Step 2. The next step in the analysis used the coefficients within each path to derive the difference between each student's expected achievement and the actual measured achievement. This was accomplished arithmetically by multiplying the student's predictor scores by the coefficients derived in Step 1 and summing to achieve the expected/typical student achievement score. This score was then subtracted from the actual achievement score to obtain the deviation score. If actual achievement for a student was higher than typical achievement for a student with that history (e.g., actual: 325; typical: 300) then the result would be positive (e.g., residual: 25). In contrast, if the actual score was less than the expected score the residual would be negative.

Step 3. The final step in the assessment was to apply Bayesian shrinkage to the result. This step is commonly used in value added analyses to reduce the impact of extreme variability across students in some teachers' classes and to account for the fact that some teachers' results are based on a relatively small number of students. To complete this step the residual data were fit as the outcome with the nesting structure illustrated in Figure 2 below.

Class composition variables were included in the HLM analysis based on the concern that peer-to-peer effects within classes had not been captured. Additionally, prior pilot data had demonstrated that models that did not include class composition effects would identify teachers whose assignments included a heavy proportion of students with disabilities as less effective than those who taught few students with disabilities. Based on prior pilot work, class composition effects were modeled at Level 2 (teacher) by the class mean prior achievement in the content area (standard deviation units), mean prior disciplinary actions, proportion of students receiving free lunch, and proportion of students diagnosed with a special education disability. Each teacher's shrunken Bayes intercept was extracted and became the student-teacher achievement outcome that was then reported back to that teacher via the CVR.

Figure 2. Two Level Model Nesting Structure of Students within Classrooms



4. Selected Results

Stability of Teacher Results across Years in Mathematics and English Language Arts

In order to examine the degree of stability of teacher outcomes across years, two sets of analyses were conducted. These analyses were conducted with the full set of data across 2007-2008, 2008-2009, and 2009-2010. It is worth noting that only a very small portion of these rosters were verified and as a result the results reported herein represent a lower bound estimate. It is anticipated that a full set of verified rosters may produce more stable results.

The first analysis examined the stability of teacher ranks across years. Within each year, teachers were ranked as having results that fell in the top or bottom 10% of teachers, top or bottom 11% to 20%, and middle 21%-80%. The data were examined for the stability of these rankings across years. The degree of stability is illustrated in Table 5 and Table 6 below.

Table 5. Stability of Teacher Ranking in Mathematics across 2008-2009 to 2009-2010

	2009-2010 Rank				
2008-2009 Rank	*Bottom 1% - 10%*	*Bottom 11% - 20%*	*Middle 21% - 80%*	*Top 81% - 90%*	*Top 91% - 99%*
Bottom 1% - 10%	26.8% (135)	18.5% (93)	46.2% (233)	4.4% (22)	4.2% (21)
Bottom 11% - 20%	14.8% (71)	15.6% (75)	62.1% (298)	5.4% (26)	2.1% (10)
Middle 21% - 80%	10.0% (508)	9.9% (504)	64.0% (3,258)	9.3% (475)	6.8% (348)
Top 81% - 90%	2.9% (14)	4.6% (22)	54.0% (259)	22.1% (106)	16.5% (79)
Top 91% - 99%	1.8% (8)	1.5% (7)	35.1% (160)	15.8% (72)	45.8% (209)

Table 6. Stability of Teacher Ranking in English Language Arts across 2008-2009 to 2009-2010

	2009-2010 Rank				
2008-2009 Rank	*Bottom 1% - 10%*	*Bottom 11% - 20%*	*Middle 21% - 80%*	*Top 81% - 90%*	*Top 91% - 99%*
Bottom 1% - 10%	22.3% (126)	17.5% (99)	52.7% (298)	4.9% (28)	2.7% (15)
Bottom 11% - 20%	17.1% (92)	15.2% (82)	59.7% (321)	5.0% (27)	3.0% (16)
Middle 21% - 80%	9.9% (575)	9.8% (566)	63.2% (3,656)	9.5% (551)	7.6% (437)
Top 81% - 90%	3.2% (17)	6.1% (33)	55.4% (298)	17.7% (95)	17.7% (95)
Top 91% - 99%	4.5% (23)	2.7% (14)	37.1% (190)	18.2% (93)	37.5% (192)

The results show moderate stability across years. Teachers who fell in the bottom 20% in 2007-2008 were likely to fall in the bottom 20% of results again (mathematics: 45.3%; ELA: 39.8. They were unlikely to move to the top of the distribution one year later. Teachers who were in the top 20% in 2008-2009 were most likely to fall in that range in 2009-2010 (mathematics: 61.6%; ELA: 55.7%). They were unlikely to move to the bottom of the distribution one year later.

Another way of examining stability is through the correlation coefficient. Table 5 and Table 6 below show the correlation coefficients between teacher results in 2007-2008, 2008-2009, and 2009-2010 relative to the number of student records available in mathematics and ELA.

Table 7. Correlation of Teacher Effects in Mathematics across 2007-2008 to 2009-2010 and 2008-2009 to 2009-2010

Minimum Number of Students Available*	**2007-2008 to 2009-2010 Correlation Coefficient** (*number of teachers*)	**2008-2009 to 2009-2010 Correlation Coefficient** (*number of teachers*)
5	.432 (3881)	.505 (4553)
10	.440 (3683)	.509 (4326)
15	.446 (3373)	.523 (3955)
20	.466 (2827)	.528 (3279)
30	.457 (2232)	.542 (2562)
40	.464 (1823)	.558 (2097)
50	.472 (1387)	.567 (1598)

* Indicates the minimum number of students available either year.

Table 8. Correlation of Teacher Effects in English Language Arts across 2007-2008 to 2009-2010 and 2008-2009 to 2009-2010

Minimum Number of Students Available*	**2007-2008 to 2009-2010 Correlation Coefficient** (*number of teachers*)	**2008-2009 to 2009-2010 Correlation Coefficient** (*number of teachers*)
5	.372 (4253)	.404 (5051)
10	.377 (4050)	.406 (4809)
15	.384 (3685)	.422 (4367)
20	.386 (3014)	.425 (3554)
30	.397 (2222)	.473 (2639)
40	.388 (1736)	.468 (2049)
50	.386 (1213)	.487 (1441)

* Indicates the minimum number of students available either year.

The data demonstrate with as few as 5 students, moderate stability was evident and that as the number of students a teacher had across two years increased, the stability increased marginally. However, the level of correlation across these two consecutive years suggests using caution in reaching conclusions from any single year's data. Further, the rank stability data in Tables 6 and 7 suggest that there is a group of teachers who will remain in the top or bottom 10% of teachers over consecutive years and about whom substantive efforts to either improve the results for their students (bottom 10%) or to retain those teachers (top 10%) may be warranted.

It is interesting to note that all of the cross-year correlations improved from the first comparison to the second. Although it is speculative at this point, it is interesting to note that the later year (2009-2010) included a substantial number of verified rosters. Perhaps increasing data quality is helping to strengthen this relationship. If that is the case, one would expect to see some additional improvement for 2009-2010 correlated with 2010-2011 and further improvement once virtually all rosters are verified.

Sensitivity of Results to Omitted Variables

Two variables, gender and ethnicity, were omitted from the pilot calculations due to the degree of social controversy surrounding their inclusion in setting expectations for teacher work and student outcomes. One group of constituents and colleagues have argued that variables such as ethnicity must be included to be fair to teachers because they are proxies for environmental advantages and disadvantages that students bring to school that are beyond teachers' control. In essence, excluding these variables will penalize the teachers of minority children if those students have achievement disadvantages that are captured by the ethnicity variable.

The alternative argument has been that it is unacceptable to include indicators for factors such as ethnicity and gender because it is unacceptable to set different expectations for students of different ethnicities. Additionally, the argument has been advanced that these variables will not contribute any meaningful information in a context with extensive prior achievement data.

To test the degree to which the inclusion of ethnicity and gender would change results, the following analyses were conducted. The models described above were rerun for mathematics and ELA with ethnicity (coded for African American, Hispanic, Asian American, or Native American) entered in one analysis and gender entered in another analysis. Tables 9 and 11, below, describe the impact of these variables on teacher outcomes.

Additionally, the impact of excluding the following variables that were included in the field test model was tested: Special Education disability, Limited English Proficiency, Section 504 status, and Free/Reduced Lunch status. Particular consideration is warranted for the special education disability and free/reduced price lunch variables. Since aggregates of these variables are included at the classroom level, both the student level and classroom aggregates were excluded when these variables were dropped from the analysis. This convention was adopted because it made little sense to include student disabilities as a classroom average, while excluding it at the student level. Tables 10 and 11 present the impact of excluding these variables on teacher outcomes.

Table 9. Impact of Adding Ethnicity or Gender to the Estimation of Teacher Effects

Content Area	Variable	Correlation	Minimum Change	Maximum Change
ELA	Ethnicity	.999	-1.66	1.81
	Gender	.998	-3.03	3.29
Math	Ethnicity	.997	-4.08	2.92
	Gender	.999	-3.89	1.20

Table 10. Impact of Removing Variables from the Estimation of Teacher Effects

Content Area	Variable	Correlation	Minimum Change	Maximum Change
ELA	Special Education*	.981	-9.37	4.31
	Limited English Proficient	.999	-2.72	3.85
	Section 504 Status	.999	-8.82	4.16
	Poverty*	.998	-2.47	2.96
Math	Special Education*	.990	-13.43	2.79
	Limited English Proficient	.999	-3.83	3.27
	Section 504 Status	.999	-4.12	1.26
	Poverty*	.999	-3.50	1.49

Table note. Variables removed at the student and teacher level simultaneously are indicated by the * character.

Table 11. Changes in Estimated Teacher Effects Resulting from Changes in Included Predictors

Content Area	**Variable**	**Percentage of Teachers with 1-2 point change**	**Percentage of Teachers with 2+ point change**
ELA	Ethnicity	0.3%	0.0%
	Gender	5.7%	0.5%
	*Special Education**	28.4%	12.7%
	Limited English Proficient	0.5%	0.3%
	Section 504 Status	2.5%	0.9%
	*Poverty**	8.5%	0.2%
Math	Ethnicity	13.5%	1.1%
	Gender	1.6%	0.3%
	*Special Education**	23.4%	6.1%
	Limited English Proficient	2.1%	0.4%
	Section 504 Status	2.9%	0.6%
	*Poverty**	1.8%	0.2%

Table note. Variables removed at the student and teacher level simultaneously are indicated by the * character. Variables whose impact was tested by removal from existing models are italicized.

Tables 9-11 require consideration of what a 1-point change in a teacher estimated effect means. One point represents 0.02 standard deviations on the re-standardized student test scores (a small difference). Generally, teacher effects fall between plus and minus 20; most teachers fall between plus and minus 10. The standard deviation of teacher effects was 9.1 for ELA and 9.8 for mathematics.

The data suggest that in the context of the prior achievement and demographic variables already included in the model, neither ethnicity nor gender substantively influence results for ELA or mathematics. Similarly, if policy makers chose to remove limited English proficiency, Section 504 status, or free/reduced lunch status, the impact on estimated teacher effects would be quite small.

The implication of removing special education disabilities information is more substantial. For some teachers, the change in estimate would be large. The proportion of teachers for whom the change will have an impact (small or large) is much greater than for any other variable considered. Finally and most importantly, the impact of excluding this variable will be highly systematic in that it will primarily impact teachers with a high proportion of students with disabilities.

Classroom Composition

The tables below describe the contribution of each classroom variable to the model. Variables were entered as the classroom mean. For categorical variables, this is the percentage of students who are members of that group.

Table 12. Level 2 Mathematics Classroom Variables for 2009-2010

Variable	Coefficient	Standard Error	T-ratio	Approximate Degrees of Freedom	P-Value
Mean Class Free Lunch	0.576	0.862	0.669	7008	0.504
Proportion of Class Special Education	-4.330	1.195	-3.623	7008	0.001
Mean Class Prior Math Achievement (SD units)	3.191	0.389	8.202	7008	< 0.001
Mean Class Suspension	-0.269	0.265	-1.016	7008	0.310

Table 13. Level 2 ELA Classroom Variables for 2009-2010

Variable	Coefficient	Standard Error	T-ratio	Approximate Degrees of Freedom	P-Value
Mean Class Free Lunch	-2.194	0.775	-2.830	7934	0.005
Proportion of Class Special Education	-4.388	0.830	-5.288	7934	< 0.001
Mean Class Prior ELA Achievement (SD units)	3.048	0.377	8.089	7934	< 0.001
Mean Class Suspension	-1.016	0.300	-3.390	7934	0.001

Across both mathematics and ELA, a striking result is that the degree to which having a high proportion of students with disabilities in a class suggests lower expected achievement for students in that class. In mathematics, a class with 100% special education enrollment would be estimated to have average achievement approximately 4.3 points lower than a class with no special education students and in ELA that estimate would be approximately 4.4 points lower. While the coefficients for prior achievement are similarly large, it is worth noting that they reflect standard deviation units (1 SD = 50 scale points). Classes whose mean achievement is a standard deviation above the mean for individuals are not common.

Estimated Average Levels of Achievement

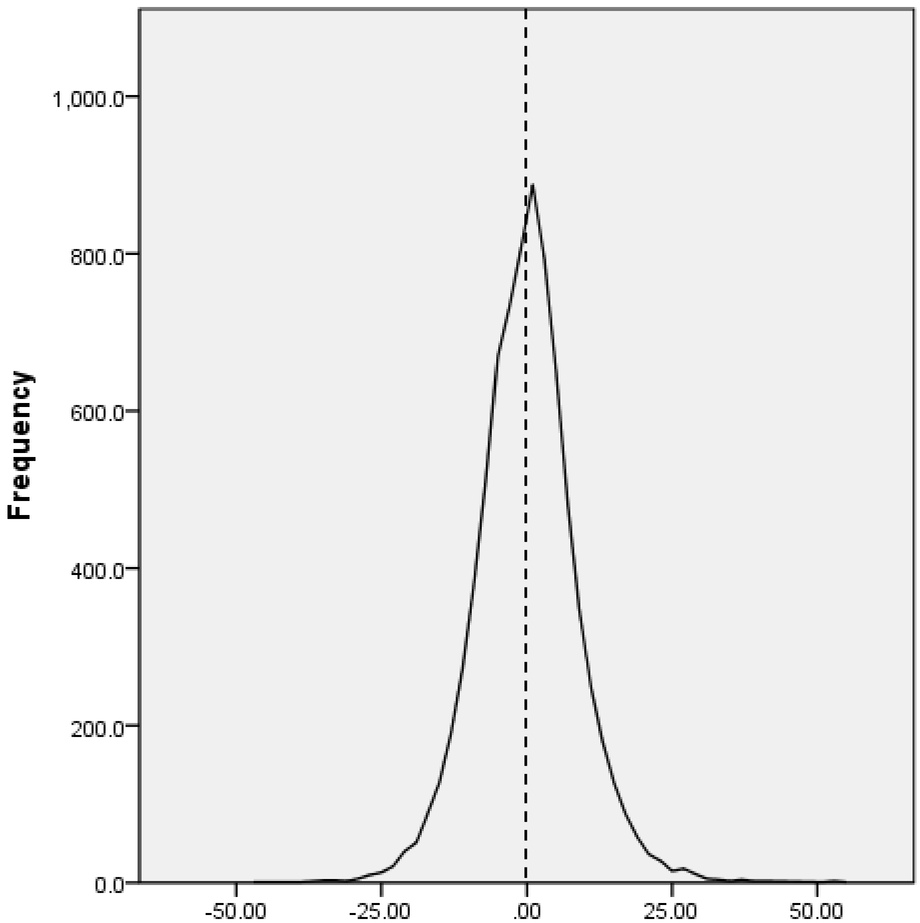
A reasoned concern that educators have expressed regarding the fairness of value added assessments is that they will not be fair because they will penalize teachers for teaching students who have historically been poorly performing. In contrast, after learning about how value added works, other teachers have expressed concern that value added will be unfair to teachers of high performing students because the more advanced the student is, the more difficult it is to make additional gains. One indicator of the extent to which these concerns emerge in the data is the correlation between the teachers' students' mean expected achievement levels and the teacher effects. If there was a substantial disadvantage in teaching historically poor performing students, there would be a positive correlation between expected achievement and teacher effects. In contrast if there was a disadvantage in teaching advanced students, there would be a negative correlation. Ideally there would be a very small to no correlation between expected achievement and teacher effects.

The data demonstrate very little correlation between predicted achievement and teacher effects for either ELA $r = 0.070$ or mathematics $r = 0.029$.

Distribution of Student-Teacher Achievement Outcomes for 2009-2010

The following figures present the distribution of outcomes across content areas for 2009-2010. The graphs depict the number of teachers (y-axis) with each magnitude of teacher effect (x-axis).

Figure 3. English-Language Arts Teacher Effects



Teacher Effect

Figure 4. Reading Teacher Effects

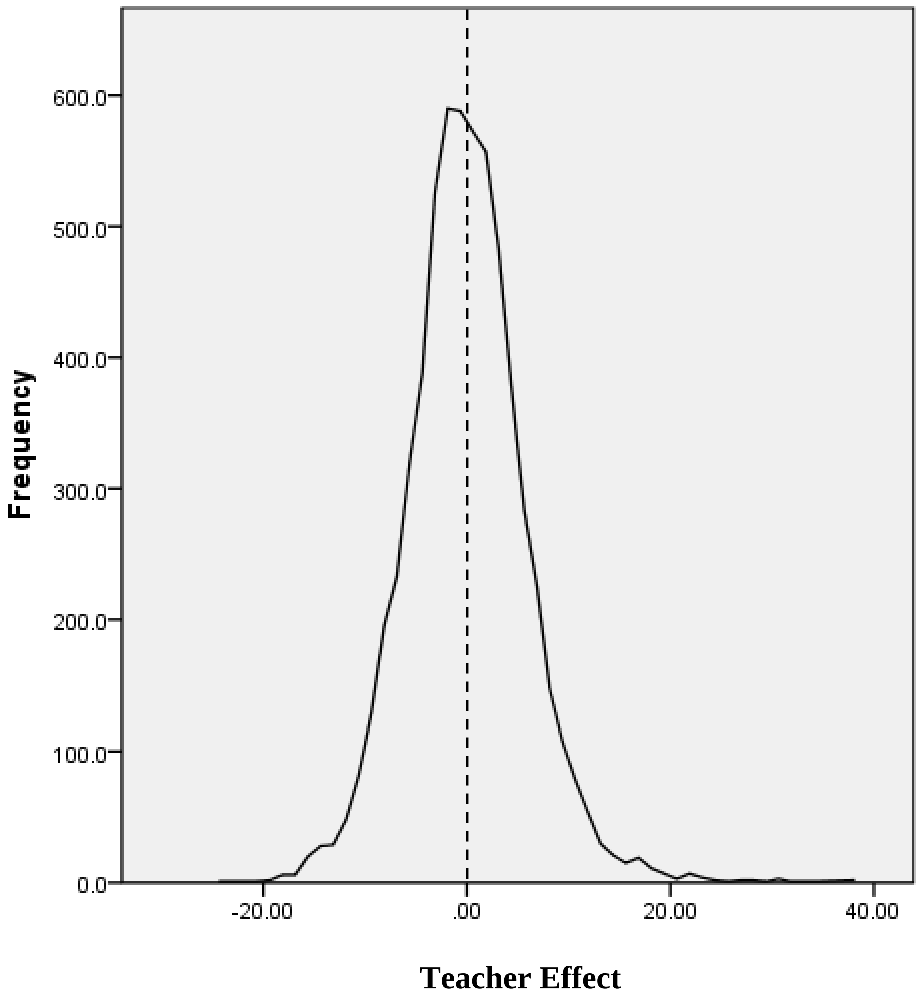


Figure 5. Mathematics Teacher Effects

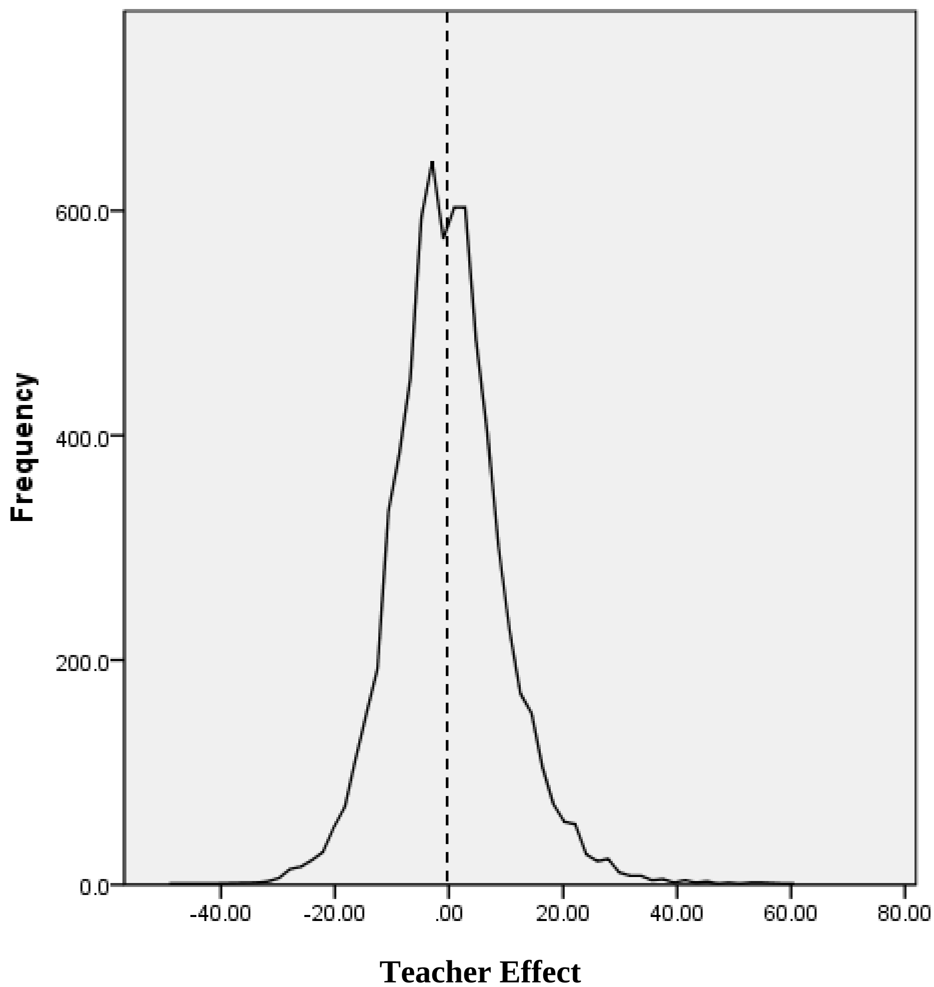


Figure 6. Science Teacher Effects

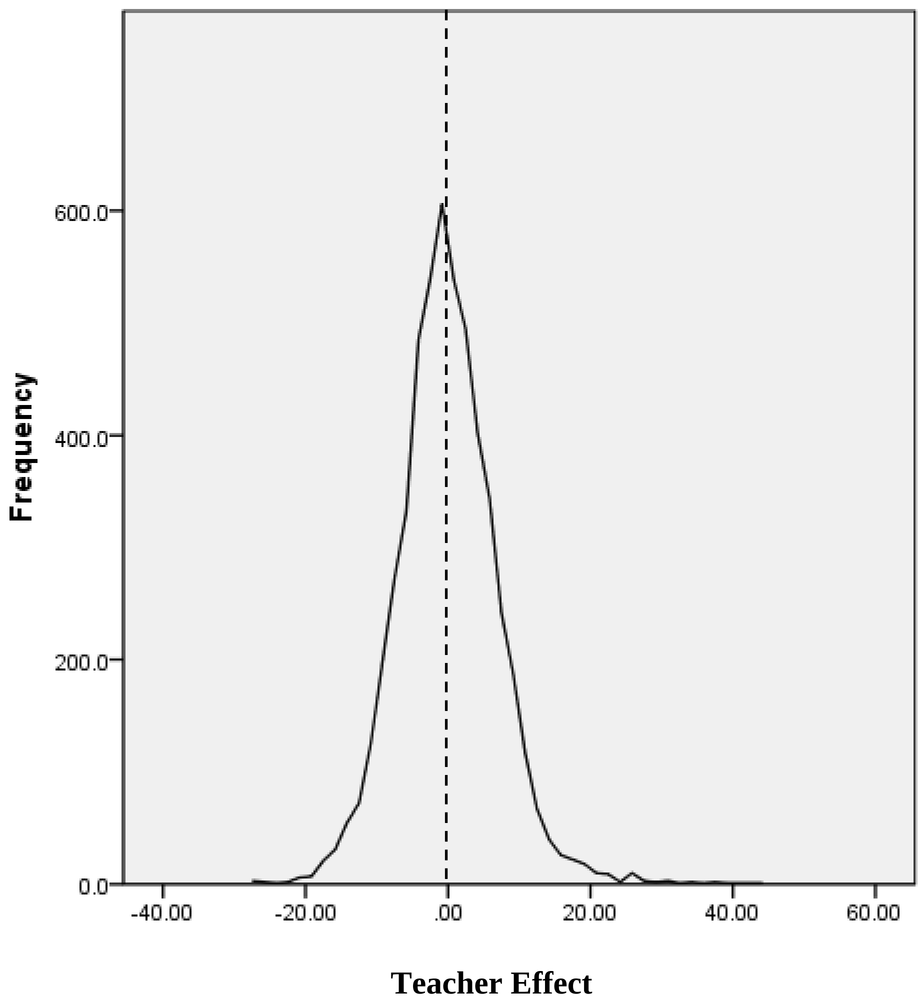


Figure 7. Social Studies Teacher Effects