Louisiana's Value-Added Assessment Model as Specified in Act 54

A Report to the Board of Elementary and Secondary Education

September 2013

Value-Added Model: Technical Process and Findings

1. Introduction

This technical brief summarizes the examination of student-teacher achievement outcomes for the 2012-2013 school year that were shared with teachers statewide in summer 2013. Outcomes were assessed via a value-added model. The assessment used regression of student data (achievement, demographics, and attendance) to estimate typical student achievement, and then compared typical outcomes to actual outcomes.

In the context of this report, *value-added analysis* (VAA) describes the use of demographics, discipline, attendance, and prior achievement history to estimate typical outcomes for students in a specific content (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 analysis 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 emerged 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). The high level of test participation in Louisiana 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 expanded to include class composition variables to address peer influences on achievement, as requested by the Advisory Committee on Educator Evaluation (ACEE).

2. Database Merging Process

Data were drawn from the standardized test files (*ITBS*, *i*LEAP, *LEAP, Algebra I EOC, and Geometry EOC*) for Spring 2010, 2011, 2012, and 2013; the Louisiana Educational Accountability Data System (LEADS) that links students to teachers; and supplemental student databases. Data analyses for 2009-2010, 2010-2011, and 2011-2012 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 or reduced price lunch status, Section 504 status, and disciplinary infractions. Data regarding teachers were drawn from the teacher demographic database (Personnel Education Profile/PEP). 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.

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 developed by the Strategic Research and Analysis (SRAA) unit at the LDOE. Details of this process are available from SRAA.

All analytic work was conducted by two Ph.D. level researchers with extensive experience with value-added models and their application to data in Louisiana, and in consultation with Dr. George Noell, Professor of Psychology at Louisiana State University.

Table 1 presents the number of records available in each content area.

Table 1. Student and Teacher Records Available Overall and in Each Content Area for 2012-2013

	Overall	**ELA grades 4-8**	**ELA grade 3**	**Mathematics Grades 4-8**	**Mathematics Grade 3**	**Science**	**Social Studies**	**Algebra I**	**Geometry**
Students	321,587	245,528	46,360	196,957	42,656	200,603	196,857	34,611	4,922
Teachers	13,866	6,263	2,364	4,821	2,164	4,230	4,527	864	395

Several important decision points are noteworthy. Initial records were limited to students who completed one assessment in grades 3-9 to permit the availability of one-year prior achievement data. The testing program begins in the 2nd grade, so, 3rd graders would be matched to 2nd grade achievement data as predictors of 3rd grade achievement. In order to be included in the analyses, a student was required to be enrolled in the same school from October 1, 2012 to March 22, 2013. These dates were set by the field test team. 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. ELA and Reading were not separated, due to the value-added model transitioning to align with Common Core standards and PARCC assessments, which do not separate ELA and reading. The ELA scaled score on both *LEAP* and *iLEAP* is derived from both ELA and Reading subtests, and teachers instructing ELA and Reading courses were eligible for the value-added analysis in ELA. Third grade students were included in the model for the first time. ELA and Math analyses were separated for 3rd grade students, due to their unique prior year achievement history including only two content areas (ELA and Math) in the 2nd grade *ITBS*. Students in grades 4-8 have a prior achievement history including four content areas (ELA, Math, Science and Social Studies) in *LEAP* and *iLEAP*. The number of records used for the analysis in Geometry was attenuated due to the fact that 10th graders were not included and 10th grade is the general admission for this test. In subsequent years, 10th graders will be included for Geometry to increase the number of records available for the analysis.

Finally, in order to be included in the analyses, students' attendance and achievement records had to be matched to the LEADS curriculum data to identify which courses 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, Algebra I, Geometry). Courses that did not fit these specific test areas, such as band, were dropped from the database.

It is important to note that full statewide deployment of the CVR occurred in three consecutive years, which allowed for comparative analyses between years. Comparative analyses between years, as described below, were based on verified rosters from the 2010-2011, 2011-2012, and 2012-2013 school years. Although, it is worth noting that participation in verification of rosters was lower in the initial pilot years, verification of rosters is increasing as more teachers and leaders become familiar with the process.

Additional work was conducted to complete the datasets. Student achievement scores were re-standardized to the mean and standard deviation across grade and promotional paths. When re-standardizing, the content scaled score was used. Promotional paths refer to how many consecutive years a student had been promoted and had predictor data (i.e., Path 3 means the student was promoted for three consecutive years; Path 2 means the student was promoted for two consecutive years, and so on). See Figure 1 for a graphical display of promotional paths.

Table 2 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 4th grade than any other grade level in the assessed span due to high stakes testing in 4th grade. Additionally, re-standardization was also required by the social context of test administration. For example, 8th 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 8th grade compared to all neighboring grades. Failure to attend to this phenomenon would result in teachers in the 7th and 9th grades being consistently found to be substantially less effective than teachers in the 8th grade, as a result of the social context of test administration.

Figure 1. Diagram of promotional paths

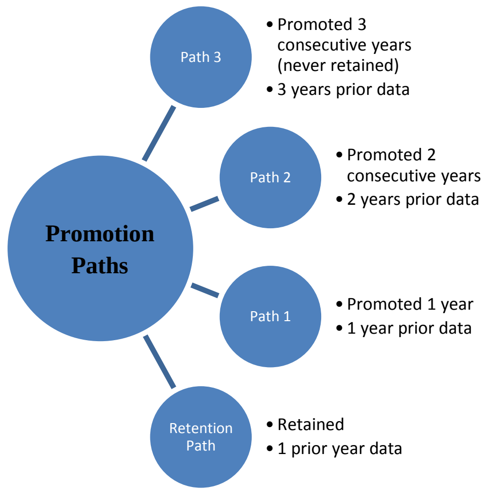

Table 2 . Number of Students in Each Promotional Path by Content Area for 2012-2013

	ELA grades 4-8	**ELA grade 3**	**Mathematics Grades 4-8**	**Mathematics Grade 3**	**Science**	**Social Studies**	**Algebra I**	**Geometry**
Path 3	123,214	N/A	93,761	N/A	98,596	96,888	30,054	3,109
Path 2	53,552	N/A	43,337	N/A	43,575	42,442	1,975	143
Path 1	61,733	46,360	54,053	42,656	52,812	52,199	2,582	915
Retention Path	6,929	N/A	5,806	N/A	5,620	5,328	N/A	N/A

Indicator variables were created to identify student characteristics. Indicator codes identified students as members of the following special education disability groups: emotional disturbance, specific learning disability, mild mental disability, speech/language impairment, other health impairment, or other special education disability. Additionally, indicator codes were used for limited English proficiency, Section 504 status, Gifted status, and free lunch and reduced lunch recipients. Indicator codes identified student characteristics using 0s and 1s. If a student has a 1 for an indicator variable, it means the student has any one of these characteristics.

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 3 displays the variables used in analyses that were included in the databases.

Table 3. Student Level Variables Examined

Variable
Emotional Disturbance
Speech and Language Impairment
Mild Mental Disability
Specific Learning Disability
Other Health Impairment
Special Education - Other
Gifted
Section 504
Free Lunch
Reduced Price Lunch
Student Absences
Suspensions (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 predictors were also entered

3. Value-Added Analysis

Once the databases were constructed, the assessment of student-teacher achievement outcomes was calculated. 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. 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. Analysis for each content area was 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 this step, data within each path were analyzed using a regression model with classroom centering to obtain the regression coefficients for each predictor. Separate intercepts were derived for each grade level.

The possibility of crossing grade by path to obtain unique grade 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 low population paths. In some atypical paths (e.g., 7th 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 typical 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 typical student achievement score. To prevent these scores from being results that are beyond the results of the assessment a capitation method was employed. The capitation method was used to lower any predicted scores that were beyond an obtainable score on the assessment. 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, as illustrated in Figure 2.

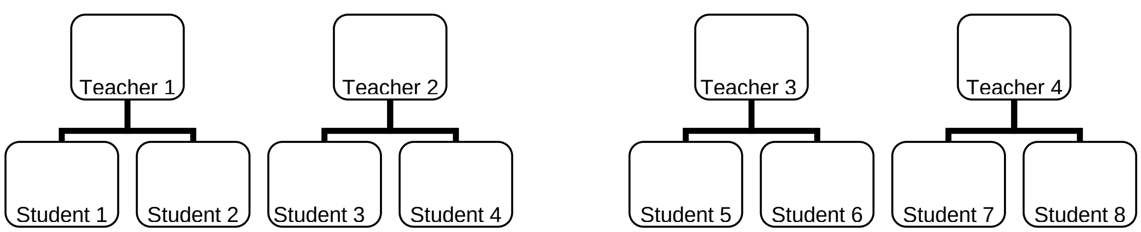

Figure 2. Two Level Model Nesting Structure of Students within Classrooms

Class composition variables were included in the Hierarchical Lineal Modeling (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. Classroom composition estimates and Bayesian shrinkage were averaged for students with multiple teachers in the same content area. Each teacher's shrunken Bayes intercept was extracted and became the student-teacher achievement outcome that was then reported to that teacher via the Compass Information System (CIS). Additionally, student-level reports were included for each teacher showing the students' expected and actual scaled scores, as well as demographic information.

Along with individual value-added scores by content, an overall composite rating was provided for the teacher. To calculate the composite percentile, the number of students a teacher instructs in each content area, along with the teacher's specific content area percentile, was compiled into one database with all teachers statewide, regardless of content. The percentile rankings for each content area were converted into a normal curve equivalent (NCE) score. A normal curve equivalent score is a score that ranges from 1 to 99 and is expressed on an equal-interval scale. This step must take place because percentiles are not on an equal-interval scale, and therefore, do not allow for arithmetic computations, such as averaging. A weighted average for the NCE provided the results for the teacher. Weighting was based on the proportion of all student results available for that teacher that each NCE represented. Once the weighted average was calculated, the NCE score was then converted back to a percentile ranking. If a teacher only teaches in one content area, that teacher's final composite percentile will not change. However, if a teacher has multiple content areas, the teacher's final composite percentile will reflect a weighted average of how he/she scored in all content areas. This composite percentile ranking will be the final value-added evaluation score that is used to determine the teacher's level of effectiveness.

4. Standards of Effectiveness

As mentioned previously, the ACEE committee was responsible for recommending standards of effectiveness for teacher evaluations. These recommendations were submitted and accepted by BESE in 2012.

For teachers where value-added data are available, the composite percentile will be converted to a 1.0-4.0 scale to use in the teacher's final evaluation. Table 4 outlines the ranges for each rating.

Table 4. Ranges for Standards of Effectiveness

Effectiveness Level	Total Score	Composite Percentile
Ineffective	1.00 - 1.45	1-10
Effective: Emerging	1.50 - 2.47	11-49
Effective: Proficient	2.50 - 3.48	50-79
Highly Effective	3.50 - 4.00	80-99

5. 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 2011-2012 and 2012-2013.

The first analysis examined the stability of teacher ranks across years. Within each year, teachers were ranked as having results that fell in the set standards of effectiveness ranges. The data were examined for the stability of these rankings across years with verified rosters. The degree of stability is illustrated in Table 5 and Table 6.

Table 5. Stability of Teacher Ranking in Mathematics across 2011-2012 to 2012-2013

	2012-2013 Rank			
2011-2012 Rank	*Bottom 1% - 10%*	*Middle 11% - 49%*	*Middle 50% - 79%*	*Top 80% - 99%*
Bottom 1% - 10% (237)	28.7% (68)	48.9% (116)	18.1% (43)	4.2% (10)
Middle 11% - 49% (1133)	13.1% (148)	48.6% (551)	29.3% (332)	9.0% (102)
Middle 50% - 79% (999)	4.8% (48)	35.0% (350)	38.1% (381)	22.0% (220)
Top 80% - 99% (669)	1.5% (10)	15.1% (101)	32.0% (214)	51.4% (344)

Table 6. Stability of Teacher Ranking in English Language Arts across 2011-2012 to 2012-2013

2011-2012 Rank	**2012-2013 Rank** *Bottom 1% - 10%*	*Middle 11% - 49%*	*Middle 50% - 79%*	*Top 80% - 99%*
Bottom 1% - 10% (267)	30.7% (82)	47.2% (126)	17.2% (46)	4.9% (13)
Middle 11% - 49% (1302)	9.2% (120)	49.9% (650)	30.3% (394)	10.6% (138)
Middle 50% - 79% (1112)	6.6% (73)	31.7% (353)	38.5% (428)	23.2% (258)
Top 80% - 99% (804)	3.1% (25)	18.0% (145)	32.8% (264)	46.0% (370)

The results show moderate stability across years. Teachers who fell in the bottom 10th percentile in 2011-2012 were likely to fall in the bottom 10th percentile of results again or to move up one ranking to the 11th - 49th percentile range (mathematics: 77.6%; ELA: 77.9%). They were unlikely to move to the top of the distribution one year later. Teachers who were in the top 10th percentile in 2011-2012 were most likely to fall in the same range or drop by one range to the 50th - 89th percentile in 2012-2013 (mathematics: 83.4%; ELA: 78.8%). 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 7 below shows the correlation coefficients between teacher results in 2010-2011, 2011-2012 and 2012-2013 school years in Mathematics and ELA.

Table 7. Correlation of Teacher Effects in Mathematics and English Language Arts across 2008-2009 to 2009-2010, 2009-2010 to 2010-2011, and 2011-2012 to 2012-2013

Content Area	**Correlation Coefficient across 2009-2010 to 2010-2011** *(number of teachers)*	**Correlation Coefficient across 2010-2011 to 2011-2012** *(number of teachers)*	**Correlation Coefficient across 2011-2012 to 2012-2013** *(number of teachers)*
Mathematics	.515 (3,948)	.504 (3,495)	.505 (3,038)
English Language Arts	.452 (4,508)	.449 (3,914)	.467 (3,485)

Estimated Average Levels of Achievement

Some educators have expressed concern regarding the fairness of value-added assessments. They have expressed the concern that value-added will not be fair because teachers will be penalized 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 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 typical 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 typical achievement and teacher effects. The data demonstrate a nearly zero correlation between typical achievement and teacher effects for either ELA (r = 0.015) or Mathematics (r = 0.045).

Distribution of Student-Teacher Achievement Outcomes for 2012-2013

The following figures present the distribution of outcomes across content areas for 2012-2013. The graphs depict the number of teachers (y-axis) with each magnitude of teacher effect (x-axis).

Figure 3. English Language Arts, Grades 4-8 Teacher Effects for 2012-2013

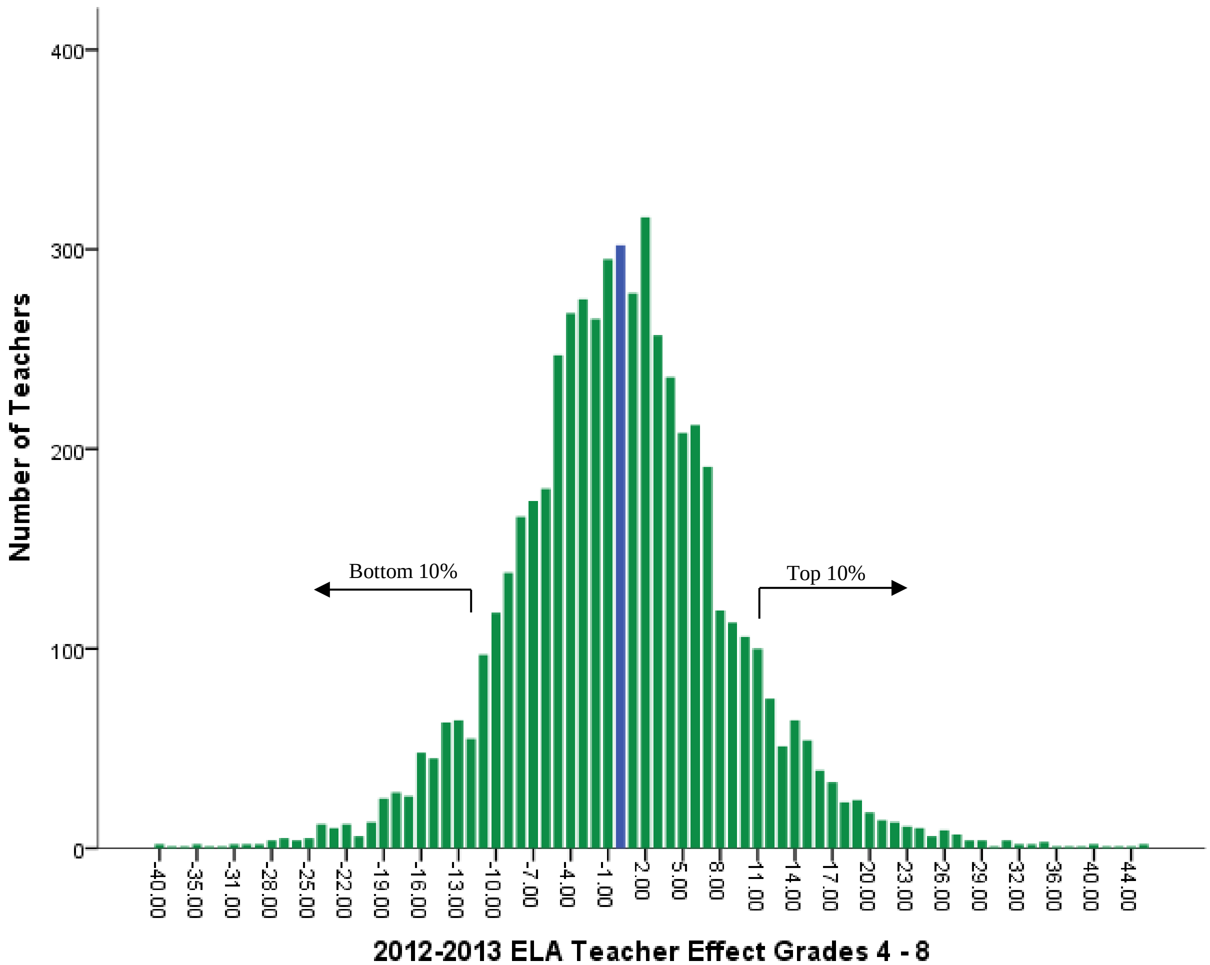

Figure 4. English Language Arts, Grade 3Teacher Effects for 2012-2013

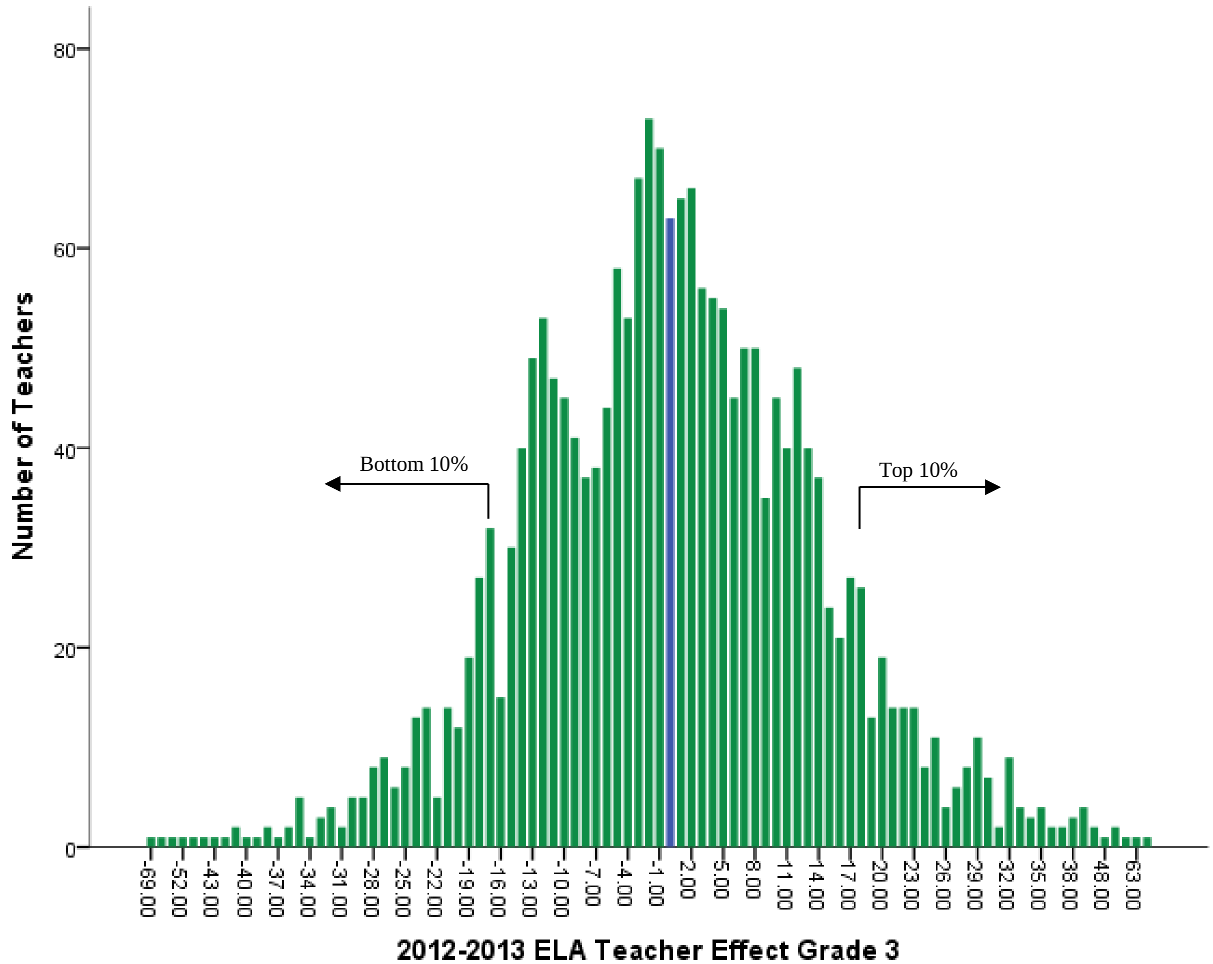

Figure 5. Mathematics, Grades 4-8 Teacher Effects for 2012-2013

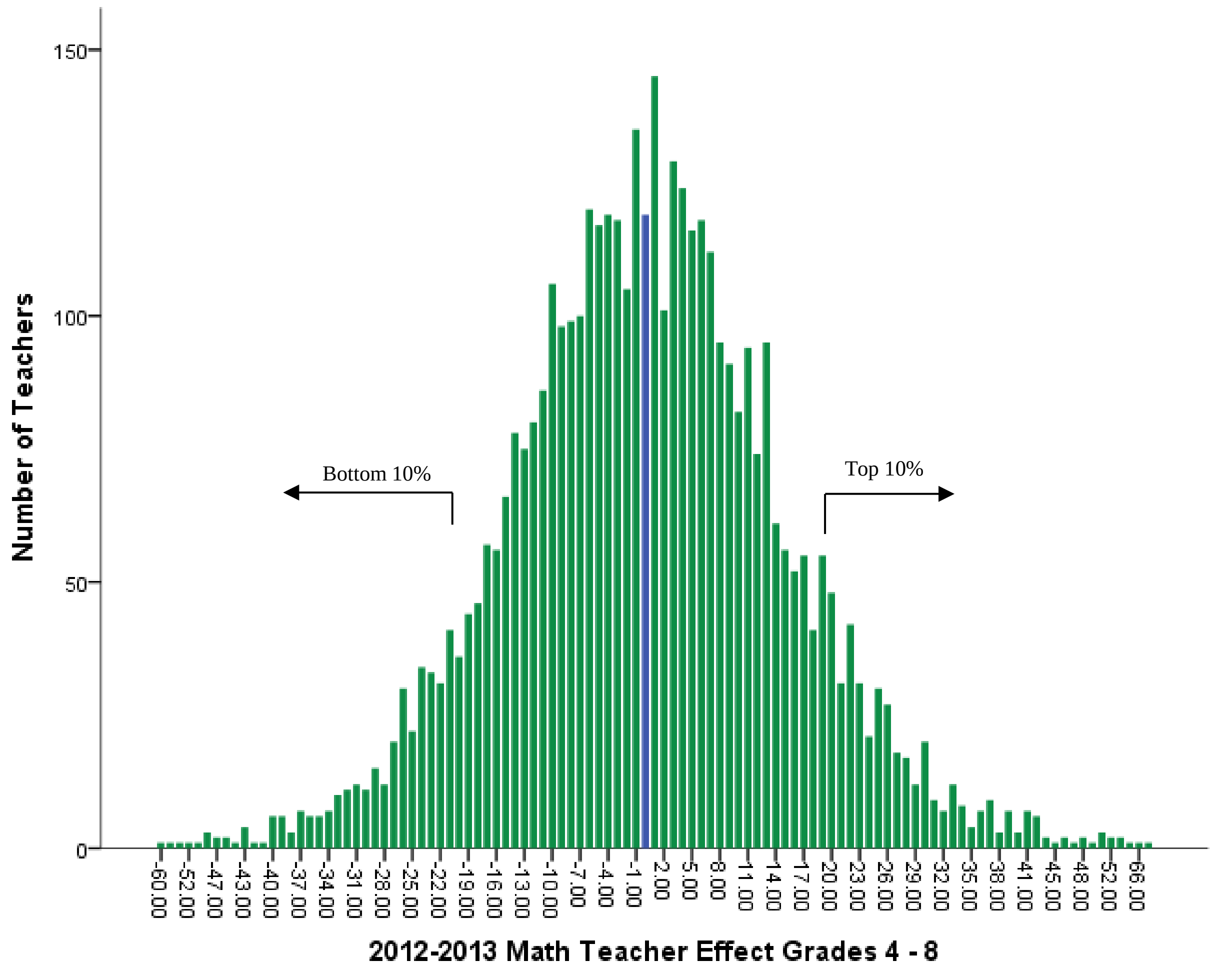

Figure 6. Mathematics, Grade 3 Teacher Effects for 2012-2013

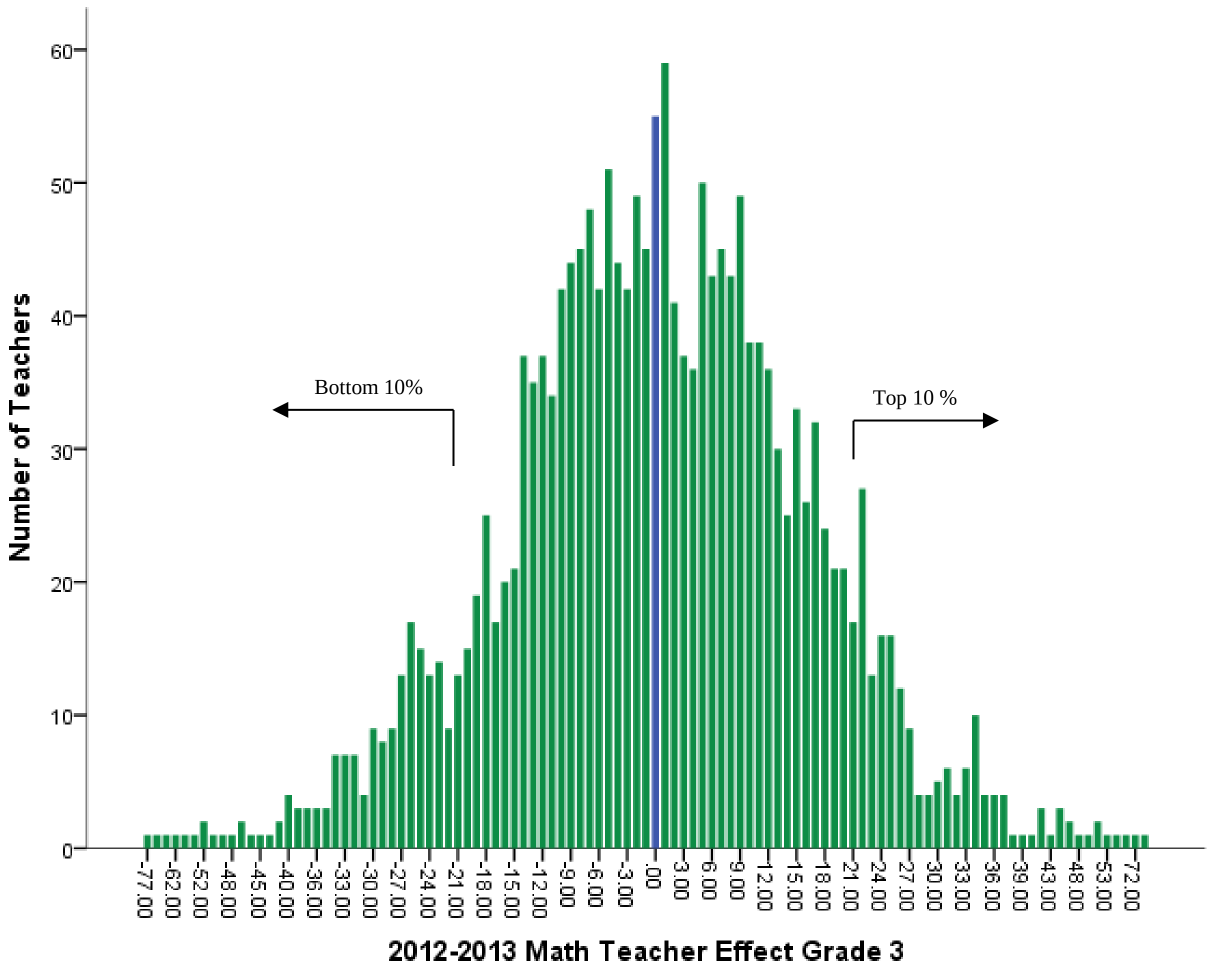

Figure 7. Science Teacher Effects for 2012-2013

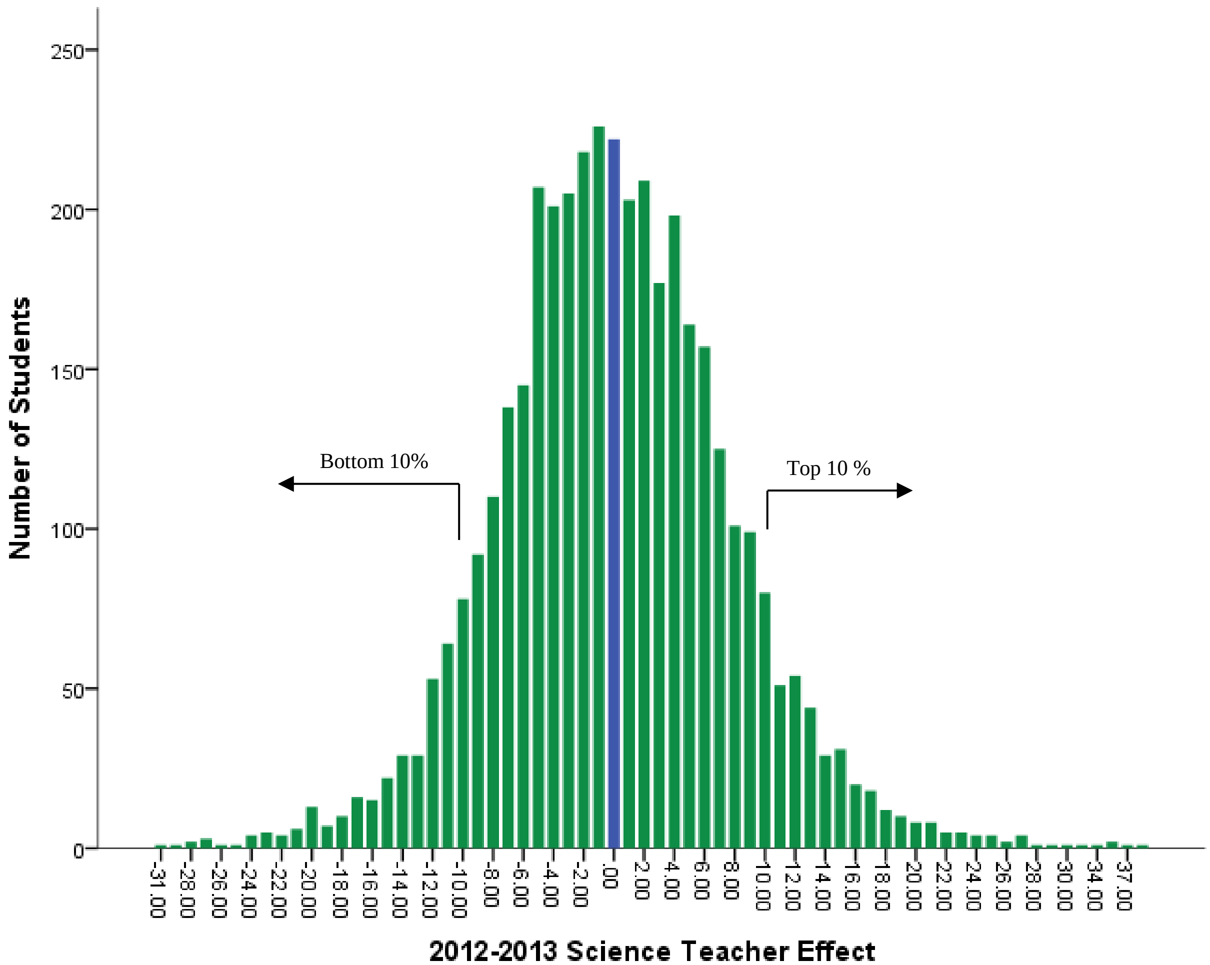

Figure 8. Social Studies Teacher Effects for 2012-2013

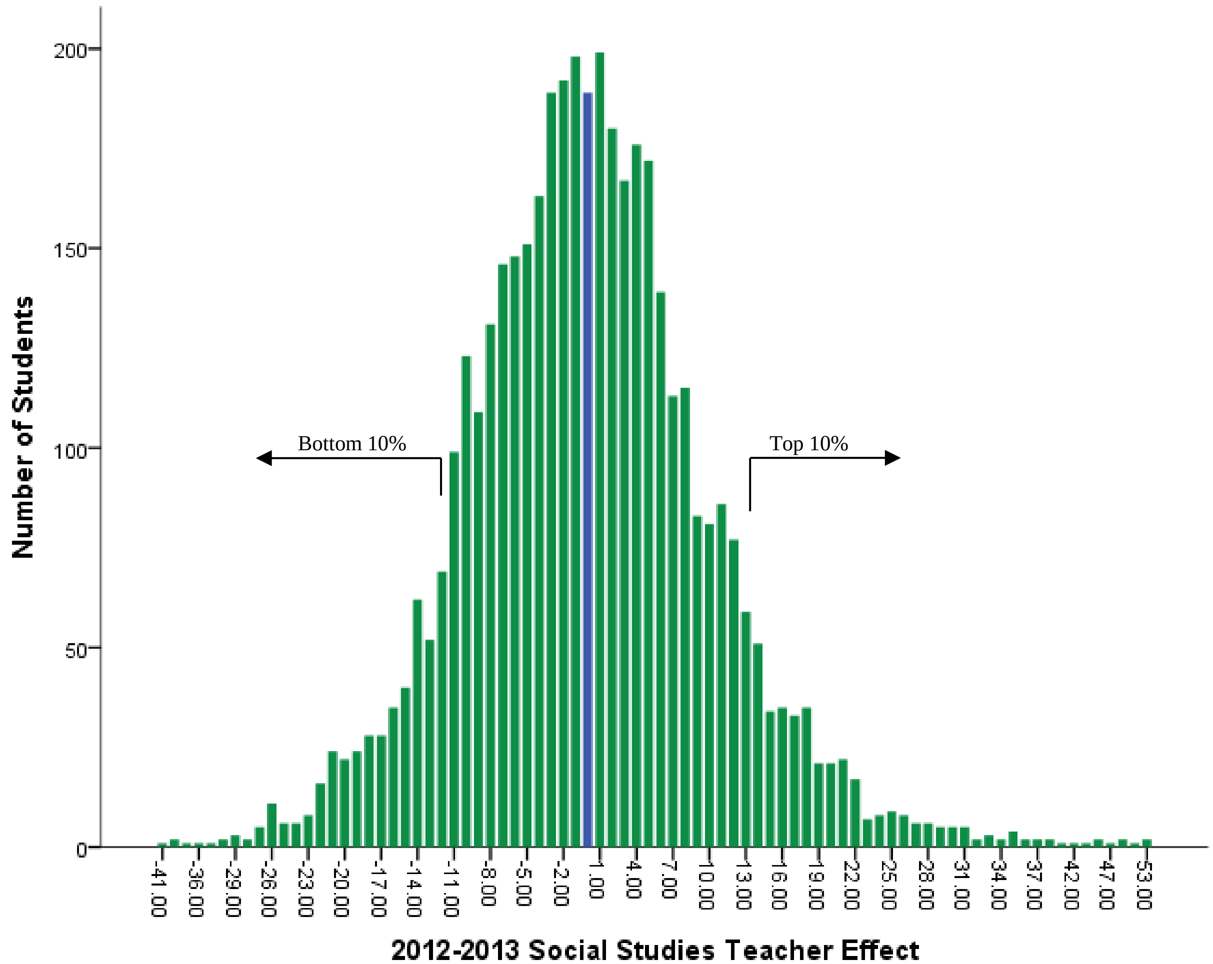

Figure 9. Algebra I Teacher Effects for 2012-2013

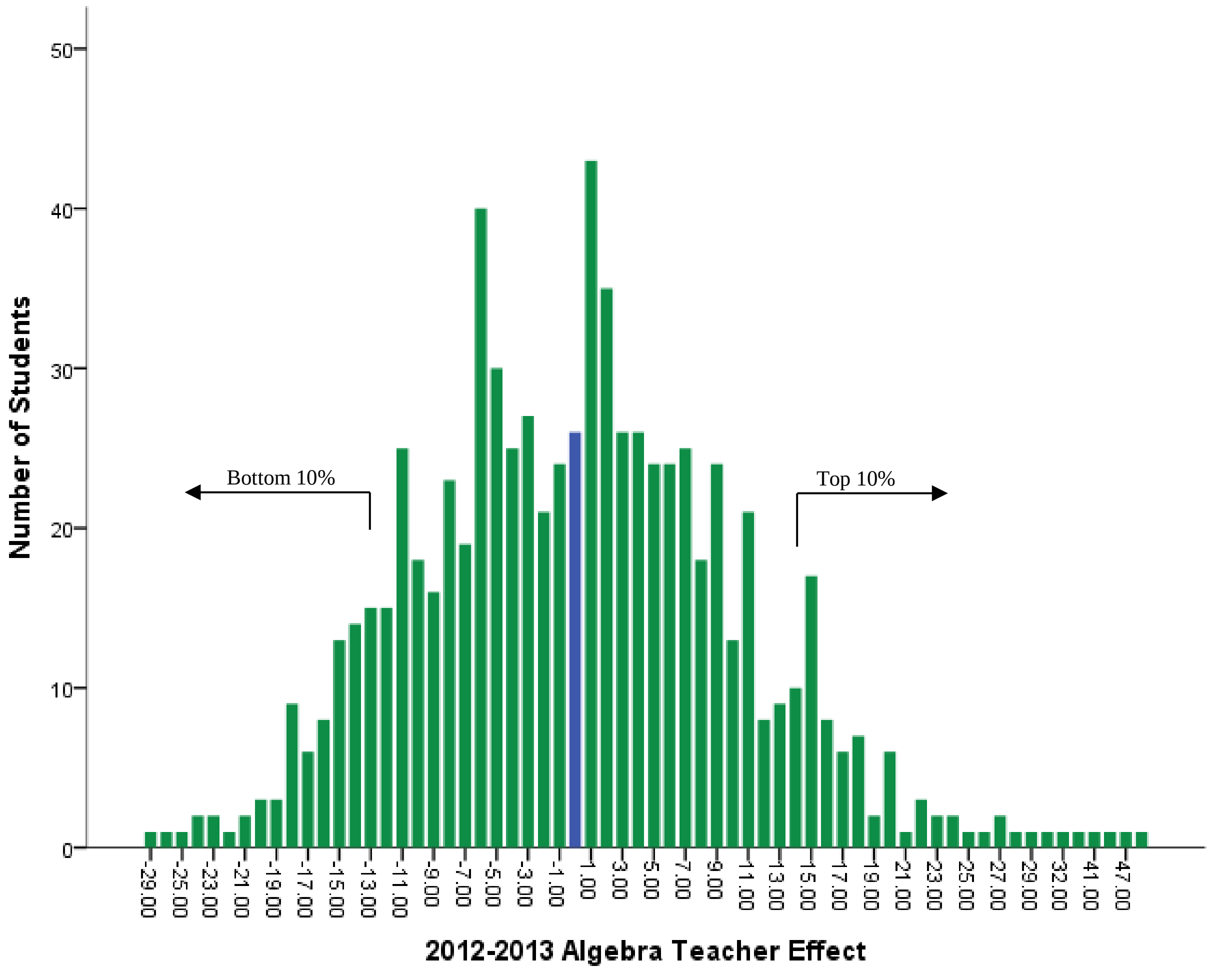

Figure 10. Geometry, Grades 6-9 Teacher Effects for 2012-2013