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**

January 12, 2012

Table of Contents

Executive Summary

Four developmental processes were deployed in support of the implementation of the value-added model required under Act 54. The first was the formation of the Advisory Committee on Educator Evaluations (ACEE). ACEE's composition included a diverse representation from across the State, with the majority of the members being practicing teachers. The second major process was the development, testing, and deployment of a secure web portal through which teachers and educational leaders were able to verify the accuracy of class rosters prior to their use in the value-added analysis, and through which they accessed their value-added reports. The third major process was the field testing of the process for providing value-added results to teachers and educational leaders. This process has been developed and refined over a two-year period. Educators have been provided with ongoing professional development and materials to prepare them to interpret their scores.

The fourth major development activity was the analytic work to prepare the results to be shared with teachers and educational leaders. The analytic work examined the impact of a number of model design choices that were reviewed by ACEE and adopted by the Board of Elementary and Secondary Education (BESE). This report provided detailed information regarding the calculation method and highlights key findings.

Notable among the findings were a group of teachers who are consistently among the teachers whose students made either the weakest or strongest educational gains. This was consistent with the results of the 2011 report. Consistent cross-year results, when they were evident for a teacher, appeared 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 was that cross-year consistency is improving as the data quality is enhanced.

I. Processes Supporting Development of the Value-Added Model

First**,** the Advisory Committee on Educator Evaluations (ACEE) was created to fulfill the requirements set forth in Act 54. The law states that at least half of the committee must be practicing educators. Of its thirty-three members, nineteen were teachers. Other panel members included parents, legislators, school board members, Board of Elementary and Secondary Education (BESE) representatives, union representatives, and other school association representatives. The committee convened its first meeting in September 2010. ACEE members were charged to make recommendations to BESE regarding the value-added model, evaluations for non-tested grades and subjects, and setting standards of effectiveness for educators. Recommendations regarding these topics were presented to BESE in December 2011.

Second, the Louisiana Department of Education (LDOE) developed and deployed the Curriculum Verification and Reporting Portal (CVR), a secure online site where teachers can verify the accuracy of their student rosters and class schedules before the data are used in the value-added assessment. The CVR was developed to address two key concerns. The first was that a number of scholars have observed that data quality was a critical barrier to accurately estimating teacher contributions to student progress and the consistency of that contribution. The second was the need to create as much transparency as practical into the process of deriving value-added scores. With the launch 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 access to the value-added results. Generally, the CVR portal is simple and follows common web conventions, with the expectation that most teachers would be able to use the portal without formal instruction. Live online training on the use of the CVR's features was provided at the request of educators. Technical support was provided for both data review and the statewide roster verification period. The portal has been tested three times (2009, 2010 and 2011), the most recent being a statewide implementation.

The third process supporting the value-added component of Act 54 has been the field testing of the educator professional development, materials and training. In 2010-2011, 19 volunteer school districts and two charter schools, for a total of 328 schools, participated in this process. The professional development included meeting with district superintendents, principals, and teacher leaders from participating schools and districts. During 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 any question that they could not answer. Professional development, ranging from one to 24 sessions, was held for each district, depending on the size of the district and the district superintendent's preferences.

The participating schools' value-added results were uploaded approximately two to three weeks following the initial training to permit remaining teachers to receive the information prior to having their scores. Follow-up meetings were held with a number of schools and districts to discuss results, concerns, and data. Feedback from these participants was collected by using a large-scale survey to determine the level of understanding of the materials, effectiveness of the trainings and materials, and to address other concerns and comments.

In summer 2011, 60 live online webinars were hosted, accessible from any location with internet access. These webinars covered a variety of topics, including general value-added information, registering and using the CVR, and reading and interpreting value-added score reports. A recorded session of each webinar topic was placed on the value-added website for continuous availability. Along with these online webinars, additional in-person training sessions have been hosted on an as needed basis and as requested basis. The LDOE team continues to attend workshops, faculty meetings, and seminars to present value-added information.

In fall 2011, schools statewide received their value-added data from the 2010-2011 school year. Trainings were available to all educators prior to the release of these scores. The fourth process supporting the 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 the LDOE staff, led by two Ph.D. level researchers with extensive experience with value-added models and their application to data in Louisiana. The remainder of this document summarizes, in brief, the analytic process and selected aggregated results from the 2010-2011 school year.

II. Technical Process and Findings

1. Introduction

This technical brief summarizes the pilot examination of student-teacher achievement outcomes for the 2010-2011 school year that were shared with teachers statewide during fall 2011. 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 ACEE.

2. Database Merging Process

Data were drawn from the standardized test files (*i*LEAP and *LEAP*) for spring 2008, 2009, 2010, and 2011; the Louisiana Educational Accountability Data System (LEADS) that links 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 or reduced price lunch status, Section 504 status, and disciplinary infractions. 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 developed by the Strategic Research and Analysis (SRAA) unit at the LDOE. Details of this process are available from SRAA by contacting Dr. Beth Gleason. 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 2010-2011

	Overall	**English Language Arts**	**Reading**	**Mathematics**	**Science**	**Social Studies**
Students	219,375	209,727	166,156	213,023	211,423	208,948
Teachers	13,189	6,532	5,631	5,660	5,031	5,426

Several important decision points are noteworthy. Initial records were limited to students who completed one assessment in grades 4-8 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 October 1, 2010 to April 4, 2011. These dates were set by the field test team. Prior to Act 54 reaching full implementation, BESE will have to set the required dates of enrollment for a student to be included in the analysis. 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. 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 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 two consecutive years, which allowed for comparative analyses between years. Comparative analyses between years, as described below, were based on verified rosters from the 2009-2010 and 2010-2011 school years. Although, it is worth noting that participation in verification of rosters was lower in the initial pilot year and incomplete in the second pilot year. 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 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 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, restandardization 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 2010-2011

	English Language Arts	**Reading**	**Mathematics**	**Science**	**Social Studies**
Path 3	96,624	67,163	98,952	97,985	97,482
Path 2	47,156	40,555	47,643	47,432	46,423
Path 1	58,174	52,202	58,628	58,409	57,479
Retention Path	7,773	6,236	7,800	7,597	7,564

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)
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 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. 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.

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. Each teacher's shrunken Bayes intercept was extracted and became the student-teacher achievement outcome that was then reported to that teacher via the CVR.

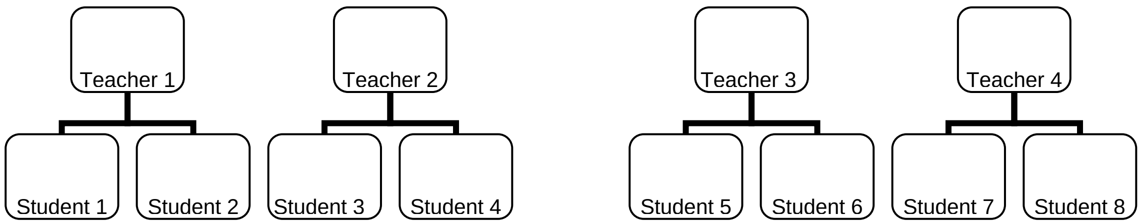

Figure 2. Two Level Model Nesting Structure of Students within Classrooms

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, were 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 December 2011.

For teachers where value-added data are available, the composite percentile will be converted to a 1.0-5.0 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	1.0 - 1.9	1-10
Effective: Emerging	2.0 - 2.6	11-25
Effective: Proficient	2.7 - 3.3	26-75
Effective: Accomplished	3.4 - 4.0	76-90
Highly Effective	4.1 - 5.0	91-99

Teachers whose value-added, composite percentile falls within the bottom 10% will receive an ineffective rating. Teachers in the middle 20-80% range will receive a rating of effective. The top 10% of teachers will receive a rating of highly effective.

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 2007-2008, 2008-2009, 2009-2010, and 2010-2011. It is worth noting that only a very small portion of the rosters were verified for the years 2008-2009, but statewide verification was implemented in the years 2009-2010 and 2010-2011.

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 with verified rosters. The degree of stability is illustrated in Table 5 and Table 6.

Table 5. Stability of Teacher Ranking in Mathematics across 2009-2010 to 2010-2011

	2010-2011 Rank				
2009-2010 Rank	*Bottom 1% - 10%*	*Bottom 11% - 20%*	*Middle 21% - 80%*	*Top 81% - 90%*	*Top 91% - 99%*
Bottom 1% - 10%	33.3% (137)	18.2% (75)	44.3% (182)	2.2% (9)	1.9% (8)
Bottom 11% - 20%	18.3% (76)	18.0% (75)	55.8% (232)	5.5% (23)	2.4% (10)
Middle 21% - 80%	7.4% (176)	9.9% (235)	67.3% (1,598)	9.1% (216)	6.4% (151)
Top 81% - 90%	3.0% (14)	3.4% (16)	57.6% (273)	17.9% (85)	18.1% (86)
Top 91% - 99%	2.0% (10)	3.2% (16)	35.5% (178)	16.6% (83)	42.7% (214)

Table 6. Stability of Teacher Ranking in English Language Arts across 2009-2010 to 2010-2011

	2010-2011 Rank				
2009-2010 Rank	*Bottom 1% - 10%*	*Bottom 11% - 20%*	*Middle 21% - 80%*	*Top 81% - 90%*	*Top 91% - 99%*
Bottom 1% - 10%	23.0% (108)	20.5% (96)	47.8% (224)	4.7% (22)	4.1% (19)
Bottom 11% - 20%	18.9% (90)	17.4% (83)	56.4% (269)	5.2% (25)	2.1% (10)
Middle 21% - 80%	7.0% (190)	10.3% (279)	66.0% (1,787)	9.7% (262)	7.0% (189)
Top 81% - 90%	4.8% (25)	5.1% (27)	56.3% (296)	17.1% (90)	16.7% (88)
Top 91% - 99%	2.8% (16)	3.2% (18)	39.5% (222)	19.9% (112)	34.5% (194)

The results show moderate stability across years. Teachers who fell in the bottom 20% in 2009-2010 were likely to fall in the bottom 20% of results again (mathematics: 51.5%; ELA: 43.5%). They were unlikely to move to the top of the distribution one year later. Teachers who were in the top 20% in 2009-2010 were most likely to fall in that range in 2010-2011 (mathematics: 59.3%; ELA: 54.4%). 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 2008-2009, 2009-2010, and 2010 2011 in mathematics and ELA.

Table 7. Correlation of Teacher Effects in Mathematics and English Language Arts across 2007-2008 to 2008-2009, 2008-2009 to 2009-2010 and 2009-2010 to 2010-2011

Content Area	**Correlation Coefficient across 2007-2008 to 2008-2009** (*number of teachers*)	**Correlation Coefficient across 2008-2009 to 2009-2010** (*number of teachers*)	**Correlation Coefficient across 2009-2010 to 2010-2011** (*number of teachers*)
Mathematics	.432 (3,881)	.507 (4,553)	.515 (3,948)
English Language Arts	.372 (4,253)	.397 (5,051)	.451 (4,508)

The data demonstrate moderate stability across years. However, the level of correlation across 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 yearly. Although it is speculative at this point, it is interesting to note that the latest years (2009-2010 and 2010-2011) included an increasing 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 2011-2012 correlated with 2010-2011, and further improvement once virtually all rosters are verified.

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 = -0.009) or mathematics (r = -0.011).

Distribution of Student-Teacher Achievement Outcomes for 2010-2011

The following figures present the distribution of outcomes across content areas for 2010-2011. The graphs depict the number of teachers (y-axis) with each magnitude of teacher effect (x-axis). The figures also display average effect for new teachers and teachers on a temporary teaching authority certificate for points of comparison.

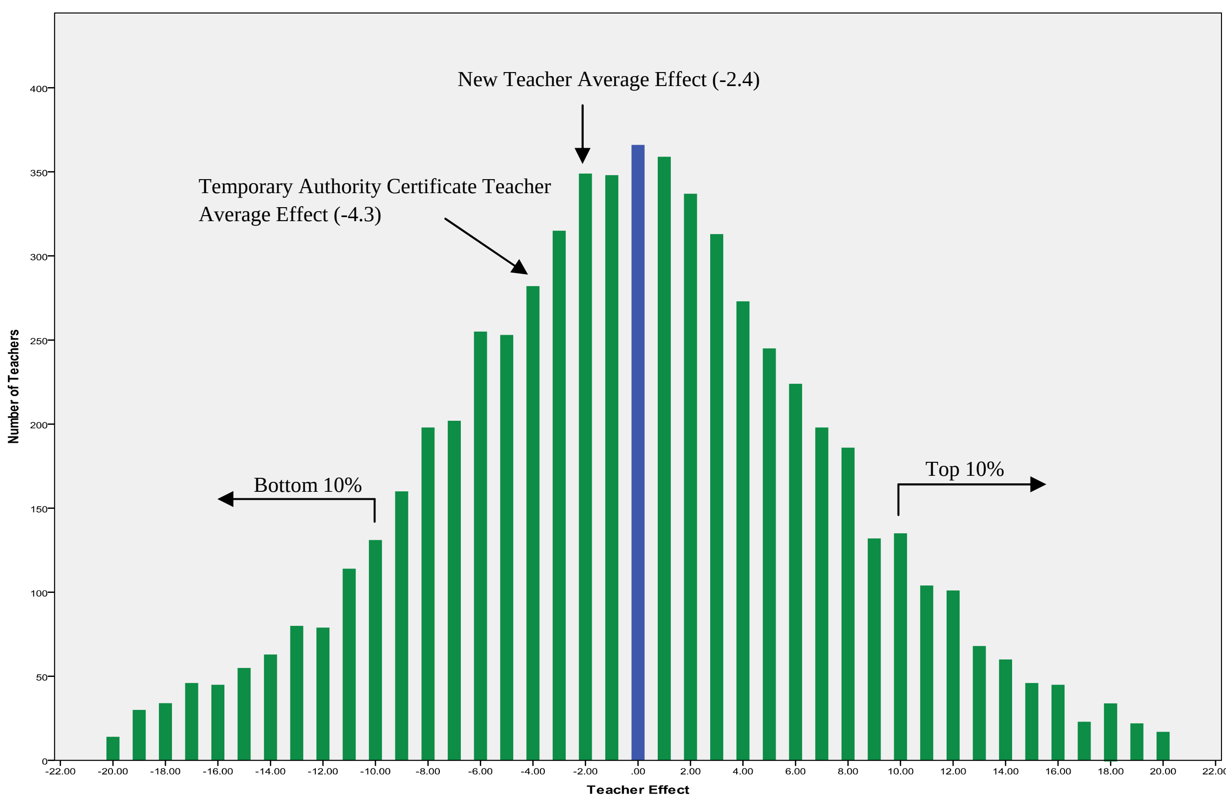

Figure 3. English Language Arts Teacher Effects for 2010-2011

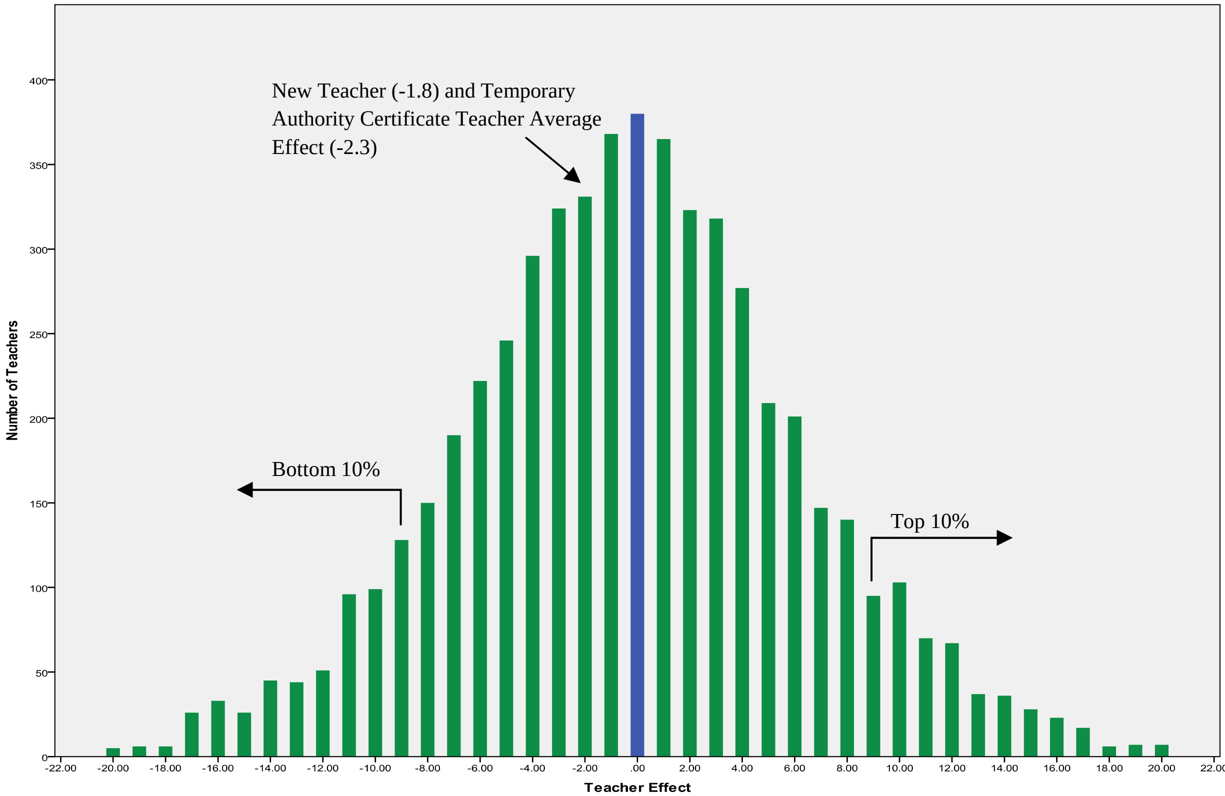

Figure 4. Reading Teacher Effects for 2010-2011

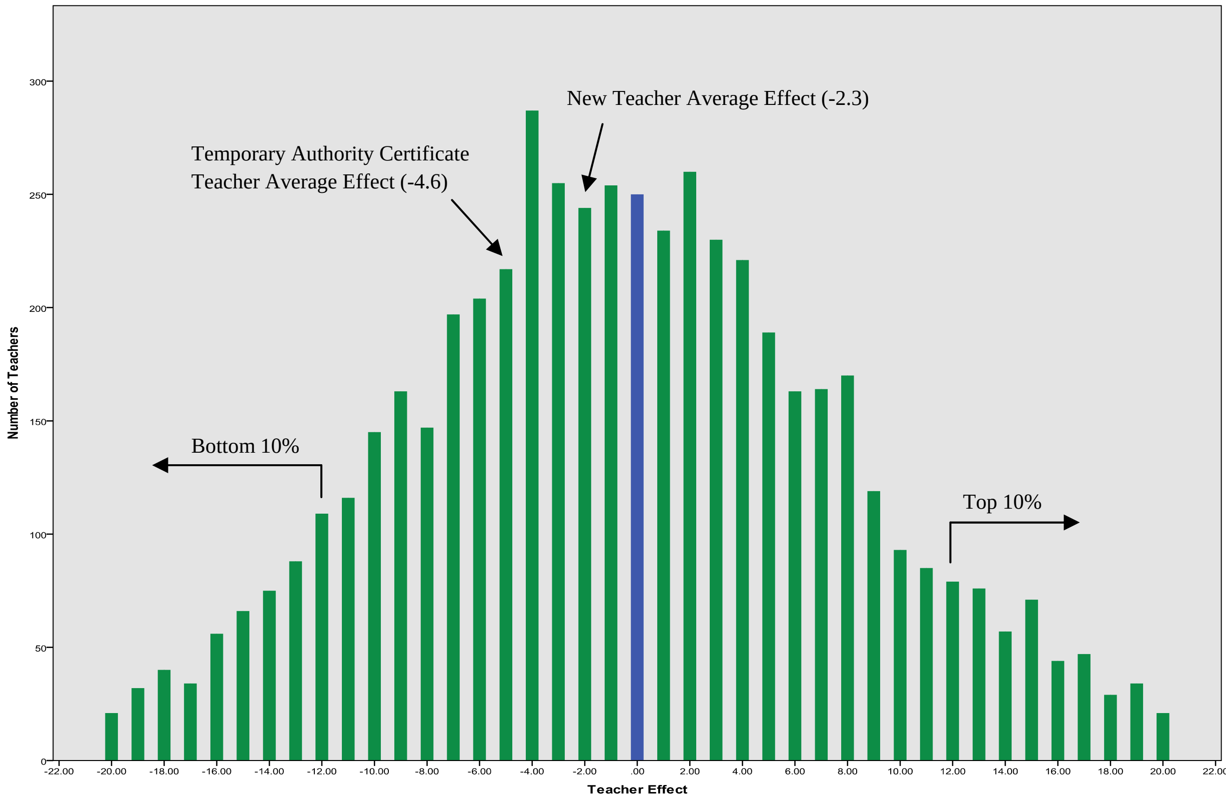

Figure 5. Mathematics Teacher Effects for 2010-2011

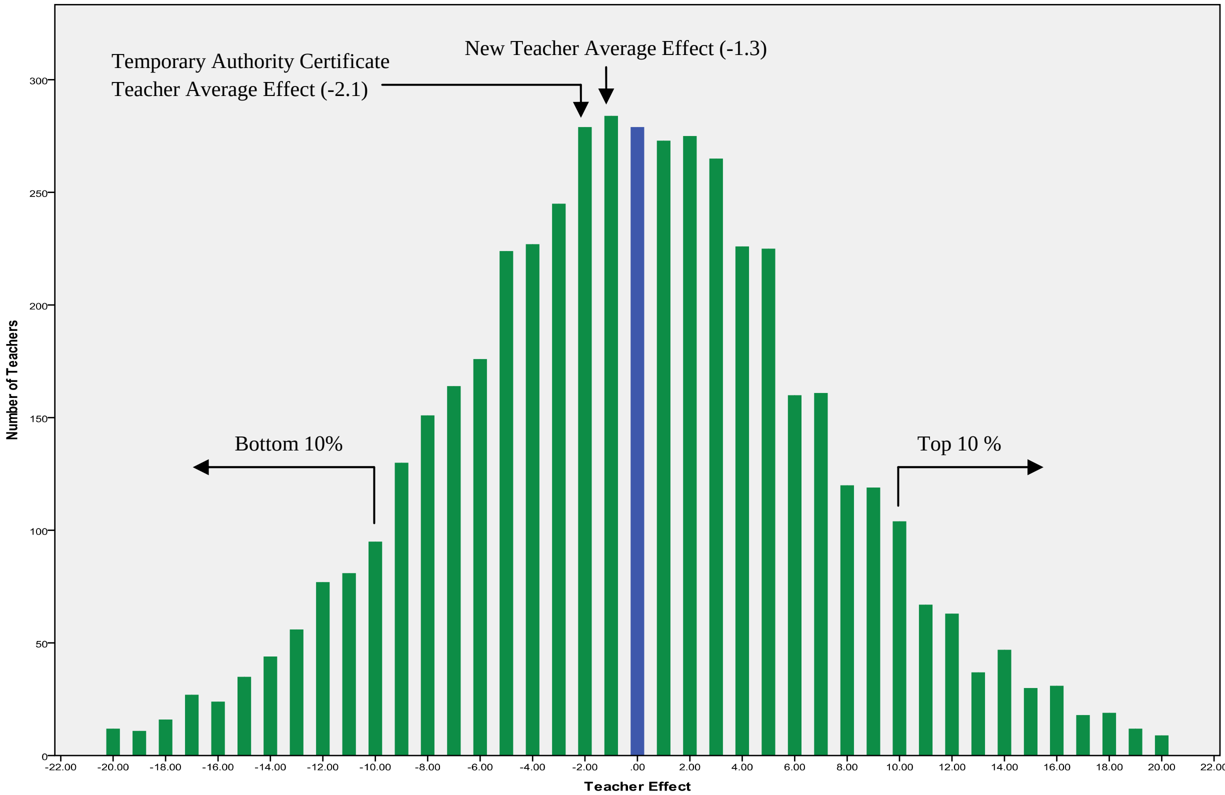

Figure 6. Science Teacher Effects for 2010-2011

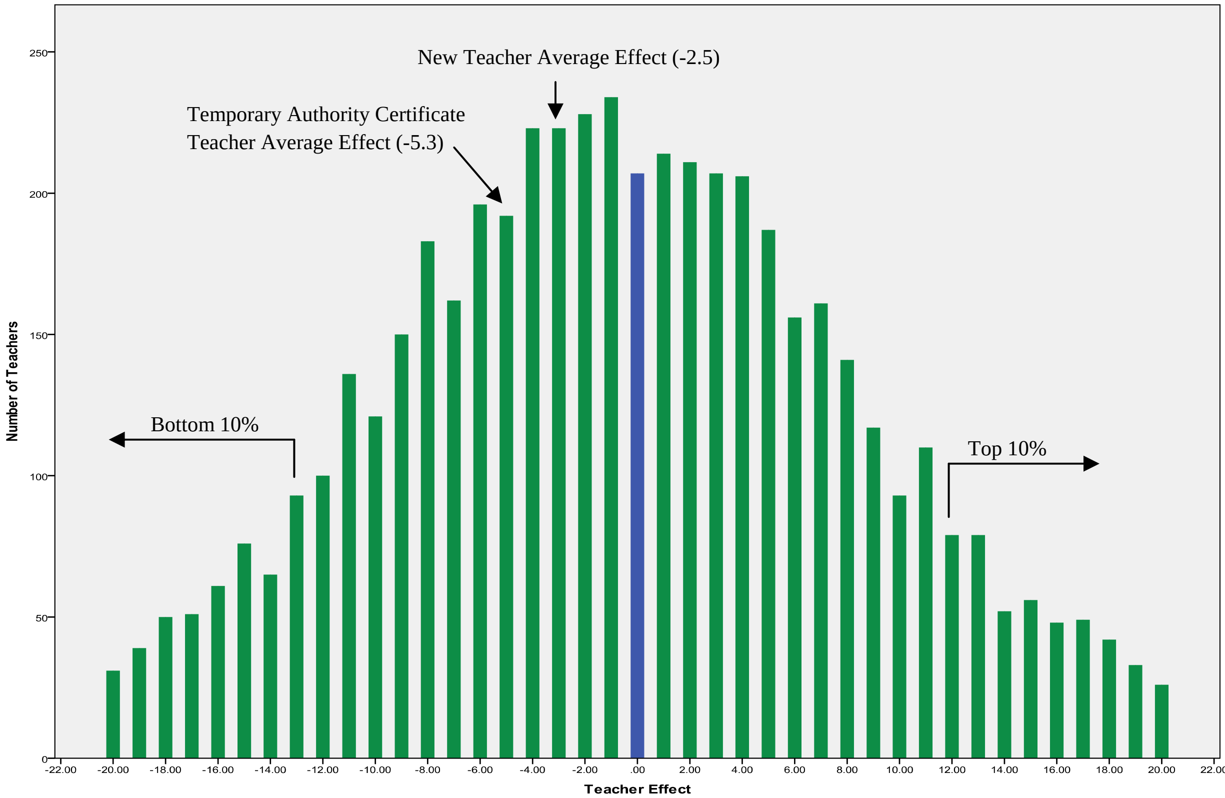

Figure 7. Social Studies Teacher Effects for 2010-2011