MATHEMATICA
Policy Research

REPORT

FINAL REPORT

Measuring Teacher and School Value Added in Oklahoma, 2013–2014 School Year

February 10, 2015

Elias Walsh
Albert Y. Liu
Dallas Dotter

Submitted to:
Office of Educator Effectiveness
Oklahoma State Department of Education
2500 N. Lincoln Boulevard
Oklahoma City, Oklahoma 73105

Submitted by:
1100 1st Street, NE
12th Floor
Washington, DC 20002
Project Director: Elias Walsh
Reference Number: 40249

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

We are grateful to the many people who contributed to this report. First, we thank the Oklahoma State Department of Education (OSDE) for funding the work. Several people at OSDE played key roles facilitating the development of the value-added model described in this report, including Kerri White, Jenyfer Glisson, Susan Pinson, and Ginger DiFalco. Jackie Skapik at Urban Policy Development also made valuable contributions. We also thank the members of the Teacher and Leader Effectiveness Commission for their guidance as the value-added model was being developed. Most of all, we thank the TLE working groups—more than 300 teachers, administrators, and education experts who dedicated many hours to building a comprehensive TLE model that meets the needs of the state of Oklahoma. The members of the Technical Advisory Board organized by OSDE—Andy Baxter, Southern Regional Education Board; Terri Grissom, Oklahoma GEAR UP; Todd Hellman, Battelle for Kids; Betina Jones-Parra, Noble High School; Daniel McCaffrey, Educational Testing Service-Carnegie Foundation; Mark Snead, RegionTrack, Inc.; and Christina Theokas, The Education Trust—also provided valuable suggestions.

At Mathematica Policy Research, Chris Rodger and Clare Wolfendale oversaw a team of programmers, including Swaati Bangalore, Lisa McCusker, Eric Lundquist, Matthew Jacobus, Alena Davidoff-Gore, and Bill Leith who processed the data and provided expert programming. Duncan Chaplin provided valuable comments. Effie Metropoulos edited the report, and Jackie McGee provided word processing and production support.

This page has been left blank for double-sided copying.

I. OVERVIEW

In this report, we describe the value-added model used as part of the state of Oklahoma's Teacher and Leader Effectiveness Evaluation System (TLE). We estimated measures of teacher and school effectiveness based on instruction provided in the 2013–2014 school year.

The value-added model described in this report is similar to the model described in the technical report for value added from the 2012–2013 school year (Walsh et al. 2014). The Oklahoma State Department of Education (OSDE) and the TLE Commission agreed to make only minor changes in how value added was calculated in the 2013–2014 school year compared to the previous year.

Teachers and administrators will receive value-added measures based on instruction provided in the 2013–2014 school year in the spring and summer of 2015. The 2014–2015 school year is the second of two pilot years for TLE's quantitative components. No stakes will be attached to the value-added results during the pilot period. The full implementation of TLE will begin in the 2015–2016 school year, incorporating value-added results based on instruction provided during the 2014–2015 school year. At that time, OSDE will combine the value-added results with additional TLE components to produce composite TLE ratings for teachers and administrators. OSDE plans for the value-added results to account for 35 percent of eligible teachers' and administrators' TLE ratings. OSDE wants educators to use the TLE ratings to promote continuous improvement of instruction and student achievement.

We worked closely with the TLE Commission and key OSDE staff to design the value-added model. The TLE Commission also sought advice from educator work groups and a six-member technical advisory board. The Oklahoma State Board of Education then made the final decisions about the model's design based on the TLE Commission's recommendations. For a broader discussion of the decision-making process and key decisions about the value-added model, please refer to the technical report from the 2012–2013 school year (Walsh et al. 2014).

In this chapter, we provide an overview of the value-added model and describe value-added methods in non-technical terms. In Chapter II, we describe how we used test scores from the 2013–2014 school year and other information about teachers and students to estimate teacher value-added results. In Chapter III, we provide the technical details of the statistical methods used to estimate value added, and in Chapter IV, we describe how we translated value-added results to the scale used in the TLE system and how we calculated value-added results for schools and student subgroups. We include tables that summarize the population of students and teachers on which the value-added estimates were based, as well as the results from the statistical model used to produce those estimates.

A. Using value added to measure teacher effectiveness

Value added is a measure of what teachers or schools contribute to students' academic growth. The measure compares the achievement of a teacher's students to an estimate of how the same students would have achieved with an average teacher. The measure is known as value added because it isolates a teacher's contribution from factors outside the teacher's control.

The basic approach of value-added models is to compare two test score averages for each teacher: (1) the average score the students actually obtained with the teacher and (2) the average estimated score the same students would have obtained with an average teacher. The difference in these two average scores—how the students actually performed with a teacher versus how they would have performed with the average Oklahoma teacher—represents a teacher's value added to student achievement. Similarly, a school's value added measures how much the school contributes to student achievement compared to what an average school contributes.

The estimated scores the students would have obtained with an average teacher—sometimes referred to as predicted scores—are called typical scores in the TLE system. OSDE chose this term because it highlights that the scores are estimated by looking at the typical achievement of students' most similar "peers" in the state—those with similar previous scores on multiple assessments and other background characteristics. Rather than comparing a student's achievement only to a relatively small number of students with identical background characteristics, we used a statistical technique called multiple regression, which simultaneously estimates a relationship between each included background characteristic and achievement. For each characteristic, this technique compares the achievement of students with the characteristic to the achievement of all other students in the state. Because a student's typical score is based on these statewide relationships between background characteristics and achievement, it represents how the student would be predicted to perform with an average Oklahoma teacher.

Because they compare actual and typical scores, value-added models enable any teacher to be identified as a high performer, regardless of the baseline achievement levels or background characteristics of the teacher's students. For example, suppose a grade 6 math teacher has a class of students who, given their background characteristics such as poverty status, disability status, and test scores on the grade 5 math, reading, and science tests (or pre-tests), typically end the year with a score of 750 on the grade 6 math test (or post-test). The value-added model calculates a relative measure of the teacher's effectiveness by comparing this class average typical score to the class average actual post-test score. In this example, if the average actual score is also 750, the value-added model will identify the teacher as an average performer because the typical and actual scores are equal. If the post-test average exceeds this standard, the teacher will be identified as above average; conversely, if the average is lower than the standard, the teacher will be considered below average.

B. The value-added models for Oklahoma

Although conceptually straightforward, value-added results are challenging to produce, as they must accurately and fairly measure the performance of teachers and schools. This requires (1) assembling an analysis file of data from multiple sources and (2) designing a value-added model that addresses Oklahoma's specific educational context. Here, we briefly describe the key elements of the analysis file (described fully in Chapter II) and then introduce the steps we used to estimate value added for teachers and schools (see Chapters III and IV for details).

We developed approaches to estimating value added based on two types of test scores from the 2013–2014 school year: (1) Oklahoma Core Curriculum Tests (OCCTs) in grades 4 through 8 in math and reading; and (2) End of Instruction (EOI) assessments for students in grades 8 and 9 for algebra I, grades 9 through 11 for geometry, grades 9 through 12 for algebra II, and grade

11 for English III. We refer to test scores from the 2013–2014 school year as post-test scores. The value-added models also use selected test scores from the 2012–2013 school year, which we refer to as pre-test scores. The value-added models yield a value-added result for each teacher on every subject they taught.

Students were eligible to be in the model if they had a post-test score and a pre-test score in the same content area from the previous grade. For example, the analysis file for students with grade 5 math post-test scores includes only those students who also have grade 4 math pre-test scores. For a student with a grade 10 geometry post-test score, the analysis file includes only students with a grade 9 pre-test score for a subject in the math content area, such as algebra I. We excluded grade repeaters so the typical scores for all students in a grade were based on a pre-test score from the previous grade in the previous year. Doing so allows for meaningful comparisons between teachers, although it does exclude some students from the value-added model. In addition to pre- and post-test scores, we collected data on other background characteristics of students, such as limited English proficiency and poverty status.

We also measured the amount of instructional time each student spent with each teacher, which we refer to as dosage. Dosage enables us to assign teachers the appropriate amount of credit for each student's performance based on two factors: (1) how much of the school year the student was in the teacher's class and (2) how much time the student spent with the teacher while enrolled. Some teachers participated in a pilot of a roster verification process in which they indicated whether and for how long they taught the students listed on their administrative rosters during each month of the school year. For teachers who participated in the pilot, we used these data to create a dosage for each teacher-student pair. However, some teachers did not teach in schools that participated in the pilot. For these teachers, we used school enrollment data to allocate proportional credit based on the fraction of time the student spent at the teacher's school.

Some students do not appear to be linked to a teacher because roster verification was not implemented statewide and because the administrative data from OSDE that linked teachers to students were limited. In these cases, we linked the student to a so-called catch-all teacher category of unassigned students for the school so these students could be included in the value-added models.

We took the following four steps to estimate the teacher value-added models. Each step addressed a different conceptual challenge.

1. **Estimating a multiple regression model.** We used multiple regression, a statistical technique that enabled us to simultaneously account for a group of background factors to avoid holding teachers accountable for factors outside their control. We accounted for a set of student characteristics that could be related to performance on the OCCT or EOI post-tests. These characteristics include pre-tests in the same content area as the post-test, pre-tests in other content areas, poverty status, gender, race or ethnicity, existence of an individualized education plan, limited English language proficiency, transfers of students between schools during the current (2013–2014) school year, and proportion of days the student attended school during the previous (2012–2013) school year. For OCCT post-test scores in math and reading, we estimated models separately for each subject and grade. For

the EOI post-test scores, we pooled eligible grades and estimated one model for each of the four subjects.

We weighted each student's contribution to a teacher's score by the proportion of time the student was assigned to the teacher while the teacher was teaching that subject. We used the Full Roster Method for teachers who shared students (Hock and Isenberg 2012). In some cases, a student was taught by one teacher for part of the year and another teacher for the rest of the year. In other cases, two or more teachers were jointly responsible for some of the same students at the same time. Using the Full Roster Method, teachers who shared students received equal credit for the students' achievement when the amount of instructional time was equal.

2. **Accounting for measurement error in the pre-test.** We used methods to account for the fact that a student's performance on a single test is an imperfect measure of ability. If we had not used these methods, teachers could be unfairly held accountable for the initial performance of their students, rather than being assessed on the gains they produce in student learning. Good or bad luck on the pre-test can dampen the observed relationship between pre- and post-test scores, compared with the true relationship between student achievement at the beginning and end of the year. If we were to use the observed relationships without making any adjustments, teachers of students with low pre-test scores might be held partly accountable for the performance of their students before they entered their classrooms. To correct for this problem, we compensated for good or bad luck in pre-test scores—also known as measurement error—by employing a statistical technique that uses data on the reliability of each OCCT and EOI test provided by the test developers.

3. **Comparing teachers across grades**. The OCCT tests are not designed to allow the comparison of scores across grades. We therefore placed teachers on a common scale by translating each teacher's value-added estimate into a metric of generalized OCCT points. We based this translation on a three-stage procedure. First, before the multiple-regression step, we translated student test scores into a common metric in which each student test score is measured relative to other test scores within the same year, grade, and subject. In doing so, we set the average student test score to zero within each year, grade, and subject. We then used these scores to produce initial teacher value-added estimates. Second, we adjusted these initial estimates so that the average teacher in each grade and subject received the same estimate. Third, we multiplied the resulting estimates by a grade-specific conversion factor to ensure that the dispersion of the estimates was similar by grade. For teachers with students in more than one grade, we took a student-weighted average of their grade-specific value-added results.

   Because the same EOI tests are given to students regardless of grade, we did not need to apply the same grade-level adjustments to the EOI value-added estimates. Instead, we accounted for grade as an additional student characteristic in the multiple regression model.[1] Although not related to grade, we applied adjustments to EOI test scores and initial value-added estimates similar to those for the OCCT model: (1) we translated student test scores into a common metric in which each student test score is measured relative to other test

[1] Although we mean-centered the student test scores and the other student characteristics in the regression, the initial value-added estimates also have to be mean-centered to account for differences in the weighting of the average due to different numbers of students contributing to each teacher's estimate.

scores within the same year and subject and (2) we adjusted these estimates so that the average teacher in each subject received the same estimate.

4. **Accounting for imprecisely estimated measures based on few students.** Value-added estimates can be misleading if they are based on too few students. Some students might score well due to good luck rather than good knowledge of the material. For teachers with many students, good and back luck affecting test performance tends to cancel out. However, a teacher with few students can receive high or low value-added results due to luck. We made two adjustments to reduce this risk: (1) we reported estimates only for teachers with at least 10 students and (2) we used a statistical technique called shrinkage that accounts for the precision of the initial value-added estimate by combining the adjusted value-added estimate (from step 3) with the overall average teacher value-added estimate to produce a final value-added result (Morris 1983). Whereas shrinkage adjusts estimates for teachers with fewer students more toward the overall average, the adjustment is smaller for teachers with many students. Thus, we relied more heavily on an assumption of average effectiveness for teachers with few students.

After estimating the teacher value-added models, we estimated a school's value-added result by averaging the teacher value-added results for teachers in the school. Consequently, a school's value added reflects the combined contributions of its teachers compared to the contributions of teachers at an average school. We gave teachers with more students more weight in the average so that students who have the same dosage in the value-added model contribute equally to the school value-added result.[2]

[2] Value-added results for the catch-all teachers of unassigned students were included in the average for schools with incomplete data linking teachers to students.

This page has been left blank for double-sided copying.

II. DATA

In this chapter, we review the data we used to generate value-added estimates of teacher and school effectiveness. First, we discuss the Oklahoma Core Curriculum Test (OCCT) and end-of-instruction (EOI) assessment scores used in the value-added model. We then discuss the data on student background characteristics used in the model. Finally, we discuss how we calculated teachers' shares of instruction when different teachers taught the same subject to the same students.

A. OCCT and EOI test scores

The outcomes we analyzed were 2014 scores from the OCCTs in math and reading and EOI assessments in algebra I, geometry, algebra II, and English III. We refer to test scores from the 2013–2014 school year as post-test scores. The value-added model also uses selected test scores from the 2012–2013 school year, referred to as pre-test scores. To be included in the value-added model, students' test score records had to meet certain conditions based on when they were tested and whether we had a record of where they were enrolled in school. The first set of conditions varies by test type:

- Students enrolled in grades 4 through 8 during the 2013–2014 school year were eligible to be included if they had an OCCT math or reading post-test score.
- Students with EOI scores in algebra I, geometry, algebra II, or English III from the 2013–2014 school year were eligible to be included if they were in a grade in which that subject area is typically taught. These grades were 8 and 9 for algebra I, 9 through 11 for geometry, 9 through 12 for algebra II, and 11 for English III.

The first two columns of Table II.1 summarize the post-test subjects and grades included in the value-added model.

We excluded test scores from the Oklahoma Modified Alternate Assessment Program (OMAAP) and OAAP assessments. We then excluded students with post-tests from the analysis file if they met any of five conditions.

1. First, we excluded students who had conflicting post-test score records for the same test type and subject, or who had scores that were not in the valid range for the test type and subject.
2. Second, we required that students were not retaking their EOI post-test; we excluded a student's post-test score for a subject if the student had an EOI score from the 2012–2013 school year for the same subject.[3]

[3] Although we excluded students who were retaking a post-test they took the year before, we did not exclude students who took the same test twice during the same school year. When a student took the same EOI post-test twice during the 2013–2014 school year, we included the post-test score from the earlier test so that all post-test scores in the value-added models reflect the scores from the first time students take a test in a subject (in the same or a different school year). When a student took the same EOI pre-test twice during the 2012–2013 school year, or took the same test in the year before, we included the pre-test score from the later test so that all pre-test scores in the value-added models reflect the most recent measure of baseline achievement in a subject.

3. Third, we excluded students who repeated or skipped a grade between the 2012–2013 and 2013–2014 school years, as they lacked pre- and post-test scores in consecutive grades and years.
4. Fourth, we required that students have a pre-test score in the same content area. For the OCCT post-tests, we required that students have pre-test scores in the same subject (math or reading). For EOI post-tests in algebra I, geometry, and algebra II, the pre-test scores must be another score in the math content area, which includes algebra I, geometry, algebra II, and OCCT math. For EOI post-tests in English III, the pre-test scores must be another score in the reading/English language arts (ELA) content area, which includes English II and OCCT reading. The third column of Table II.1 lists the possible same-content pre-tests for each post-test subject.
5. Finally, we excluded students from the analysis file if they were not linked to a teacher eligible to be in the value-added model. To be eligible, a teacher had to teach at least five students in his or her grade for an OCCT subject, or across eligible grades for an EOI subject. We do not estimate a value-added measure for teachers with so few students.

Some students were not linked to any teacher in a subject. An unassigned student that was not excluded for any of the five reasons above can still be linked to a so-called catch-all teacher category for the student's school. A catch-all teacher stands in for the set of teachers in a school with missing or incomplete links to students in the data we received from OSDE. In doing so, we included an unassigned student in the value-added model if the student's catch-all teacher meets the same five-student eligibility requirement for an OCCT or EOI subject.[4]

Table II.1. Value-added model test subjects and grades

Post-test subject	Post-test grades	Same-content pre-test subjects
OCCT math	4 through 8	OCCT math
OCCT reading	4 through 8	OCCT reading
Algebra I EOI	8 and 9	OCCT math
Geometry EOI	9 through 11	OCCT math, algebra I, algebra II
Algebra II EOI	9 through 12	Algebra I, geometry
English III EOI	11	English II

Note: For a post-test score to be included in the value-added model, the student must have a pre-test score from the same content area in the previous grade.

After applying these rules, we reported estimates only for teachers who taught 10 or more students over the course of the 2013–2014 school year in at least one subject in any grade. Only students who were linked to a teacher who met the five-student threshold and were included in the analysis file counted towards this 10-student minimum. For example, we would report an estimate in reading for a teacher who claimed seven students in reading in grade 4 and six students in grade 5. For a teacher who claimed nine students in reading in grade 4 and four

[4] We included a catch-all teacher for each grade in a school with unassigned students in the value-added models for OCCT math and reading, and a single catch-all teacher that captured unassigned students in any grade in a school in the value-added models for EOI subjects.

students in grade 5, however, the grade 5 students would not be linked to the teacher, as they would not meet the five-student minimum in that grade level. Because such a teacher would be linked to only nine students across all grades, we would not report a value-added estimate in reading for this teacher.

Table II.2 shows the total number of students who could have been included in the analyses, the reasons why excluded students were removed, and the total number of students whose test results were used in the models after the exclusions had been made. The second through fifth columns show the totals for students in OCCT math and reading, and the last two columns show the totals for all four EOI subjects combined. The top row shows the total number of students with post-test scores. As shown in the bottom row of the table, 85.0 percent of students with test scores from 2013–2014 were included in the analysis file for OCCT math, 85.6 percent for reading, and 69.0 percent for EOI subjects. The most common reason students were excluded was for missing a pre-test score from the same content area as the post-test.

The pre-test subjects associated with each EOI post-test subject vary because not all students take the EOI test in a subject in the same grade, and because there is no set order in which students must take the courses associated with the tests. We show the distribution of eligible pre-tests for each EOI post-test in Table II.3. For example, 74.3 percent of students with algebra I post-test scores in the analysis file have grade 8 OCCT math pre-tests, and 25.7 percent have grade 7 OCCT math pre-tests. These percentages sum to 100 percent because we required that all students in the analysis file have a pre-test score in the same content area.[5] In contrast, only 98.1 percent of students with an algebra I post-test score have a pre-test score in the reading/ELA content area because we do not require that algebra I students have a reading/ELA pre-test.

The OCCT and EOI scores ranged from 400 to 999. However, OCCT scores on these scales are not designed to be meaningfully compared between grades, years, and subjects, nor can EOI scores be compared meaningfully between years and subjects. To compare OCCT test scores across grades within each subject and year, we transformed the test scores in a two-part process. First, we subtracted the mean score and divided by the standard deviation for each grade, subject, and year to obtain a *z*-score.[6] This step enabled us to translate the math and reading content scores in every grade into a common metric. Second, we created a measure with a range resembling the original test score scale by multiplying each *z*-score by a common factor across all grades within each subject and year. The common factor was equal to the square root of the average variance across all grades for each subject and year. For EOI scores, we subtracted the mean score for each subject, and year, but did not change the standard deviation.

[5] For students with scores from two or more different math assessments from the 2012–2013 school year, we included only the score on the higher-level assessment. We considered OCCT assessments as lower than EOI assessments. Algebra II is the highest-level EOI assessment, and geometry is higher than algebra I. For example, only algebra I scores were used as pre-tests for students who took both the grade 8 math and algebra I assessments in 2012–2013.

[6] Subtracting the mean score for each subject and grade creates a score with a mean of zero in all subject-grade combinations.

Table II.2. Reasons that students tested in 2014 were excluded from the analysis files

	OCCT math		OCCT reading		EOI subjects	
	Number	Percentage	Number	Percentage	Number	Percentage
Students with post-test scores	229,064	100.0	239,519	100.0	168,818	100.0
(1) Conflicting post-test scores	57	0.0	59	0.0	1,320	0.8
(2) Score in same subject as EOI post-test from previous year	0	0.0	0	0.0	15,235	9.0
(3) Skipped or repeated a grade	1,604	0.7	1,629	0.7	2,888	1.7
(4) Missing pre-test score from same content area	29,002	12.7	30,640	12.8	30,360	18.0
(5) Not linked to an eligible teacher	3,609	1.6	2,247	0.9	2,478	1.5
Total excluded	**34,272**	**15.0**	**34,575**	**14.4**	**52,281**	**31.0**
Total included	**194,792**	**85.0**	**204,944**	**85.6**	**116,537**	**69.0**

Source: OSDE administrative data.

Notes: The table does not include 7,936 student-subject combinations with OMAAP or OAAP scores from 2014, but no OCCT or EOI tests. The table includes only students in eligible grades for a post-test subject.

Students are excluded sequentially in the order presented and so do not count for more than one reason in this table.

The columns for math and reading include students in grades 4 through 8. The EOI subjects are algebra I, geometry, algebra II, and English III.

For OCCT math, algebra I, geometry, or algebra II, the same-content pre-test score is another mathematics assessment. For OCCT reading or English III, the same-content pre-test score is another reading/ELA assessment.

For OCCT subjects, teachers must be linked to at least five eligible students in a single grade level to be considered eligible to be included in the value-added model. For EOI subjects, teachers must be linked to at least five eligible students in any grade to be considered eligible. We then reported estimates in a subject only for teachers who taught 10 or more students in any grade. The table counts students linked to teachers who do not meet the 5-student threshold as excluded for not being linked to an eligible teacher. Because we estimated (but did not report) value-added results for teachers who did not meet the 10-student threshold, the table counts students linked to those teachers as included in the analysis files.

Post-tested students not linked to any teacher in a subject are linked to catch-all teachers of unassigned students for the school-grade combination. These catch-all teachers are considered eligible teachers if they are linked to at least five eligible students.

Table II.3. Pre-test subjects of students by EOI post-test subject

	Post-test subject							
	Algebra I		Geometry		Algebra II		English III	
Pre-test subject	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Math								
Grade 7 OCCT	8,363	25.7	0	0.0	0	0.0	0	0.0
Grade 8 OCCT	24,118	74.3	311	1.0	0	0.0	0	0.0
Algebra I	0	0.0	28,253	91.2	3,874	14.9	2,801	10.4
Geometry	0	0.0	0	0.0	22,209	85.1	15,420	57.1
Algebra II	0	0.0	2,399	7.7	0	0.0	6,979	25.8
No math pre-test	0	0.0	0	0.0	0	0.0	1,810	6.7
Total	**32,481**	**100.0**	**30,963**	**100.0**	**26,083**	**100.0**	**27,010**	**100.0**
Reading/ELA								
Grade 7 OCCT	8,337	25.7	0	0.0	0	0.0	0	0.0
Grade 8 OCCT	23,534	72.5	7,167	23.1	0	0.0	0	0.0
English II	0	0.0	3,876	12.5	14,151	54.3	27,010	100.0
No reading/ELA pre-test	610	1.9	19,920	64.3	11,932	45.7	0	0.0
Total	**32,481**	**100.0**	**30,963**	**100.0**	**26,083**	**100.0**	**27,010**	**100.0**
Science								
Grade 8 OCCT	23,911	73.6	7,193	23.2	0	0.0	0	0.0
Biology I	0	0.0	9,222	29.8	12,900	49.5	14,969	55.4
No science pre-test	8,570	26.4	14,548	47.0	13,183	50.5	12,041	44.6
Total	**32,481**	**100.0**	**30,963**	**100.0**	**26,083**	**100.0**	**27,010**	**100.0**

Source: OSDE administrative data.

Notes: All percentages are based on the total count for the post-test subject.

To be included in the model, students are required to have a pre-test score in the same content area as the post-test. For OCCT math, algebra I, geometry, or algebra II, the same-content pre-test score is another mathematics assessment. For OCCT reading or English III, the same-content pre-test score is another reading/ELA assessment. Students are not required to have pre-tests in other content areas.

B. Student background characteristics

We used the data provided by OSDE to construct variables used as controls for student background characteristics in the value-added model. The value-added model accounts for the following:

- Prior achievement in the same content area as the post-test
- Prior achievement in other content areas (including math, reading/ELA, and science, when available)
- Poverty status
- Gender
- Race/ethnicity
- Existence of an individualized education plan (IEP)
- Limited English language proficiency
- Transfers of students across schools during the 2013–2014 school year
- Proportion of days the student attended school during the 2012–2013 school year

Attendance is a measure of student motivation. We used student attendance the year before—rather than the current year—to avoid confounding student attendance with current-year teacher effectiveness; that is, a good teacher might be expected to motivate students to attend school more regularly than a weaker teacher would. The proportion of the days a student attended school is a continuous variable that could range from zero to one. Because some districts did not provide OSDE with student-level attendance records from the 2012–2013 school year, we used the typical attendance rate from the student's school and grade in place of the student's individual attendance rate for 44.5 percent of students in the analysis files.

Aside from attendance and pre-test variables, the student background variables are binary, taking a value of zero or one. In the OCCT value-added model we accounted for the existence of an IEP and limited English proficiency separately for students with and without accommodations on the post-test assessment. In the EOI models we pooled students with IEPs and students with limited English proficiency into single categories because too few students received accommodations to include these students in separate categories. Table II.4 shows the characteristics of students in the OCCT and EOI analysis files.

We imputed data for students who were included in the analysis file, but had missing values for one or more student characteristics. Our imputation approach used the values of non-missing student characteristics to predict the value of the missing characteristic. Less than 3 percent of students in the value-added analysis files had any characteristic imputed. Most imputed values were for missing pre-test scores in different content areas from the OCCT post-test.[7] We did not generate imputed values for the same-content pre-test; rather, we dropped from the analysis file

[7] In addition to imputing values for some of the characteristics included in Table II.4, we also generated imputed values of attendance during the 2012–2013 school year for less than 1 percent of students.

any students with missing same-content pre-test scores. Finally, we did not impute any missing pre-test scores for students in the EOI analysis file.

Table II.4. Characteristics of students from the 2013–2014 school year

Characteristic	OCCT math Number	OCCT math Percentage	OCCT reading Number	OCCT reading Percentage	EOI subjects Number	EOI subjects Percentage
Included in the value-added model	194,792	100.0	204,944	100.0	116,537	100.0
Eligible for free lunch	95,883	49.2	97,927	47.8	41,944	36.0
Eligible for reduced-price lunch	17,973	9.2	18,763	9.2	10,179	8.7
Female	97,279	49.9	103,015	50.3	60,049	51.5
African American	22,622	11.6	23,470	11.5	12,196	10.5
Hispanic	29,433	15.1	30,129	14.7	14,485	12.4
American Indian	45,381	23.3	47,159	23.0	25,887	22.2
Asian/Pacific Islander	5,458	2.8	6,118	3.0	3,392	2.9
Caucasian/other	138,284	71.0	145,935	71.2	83,506	71.7
Individualized education plan with accommodations	9,534	4.9	10,413	5.1	n.a	
Individualized education plan without accommodations	12,066	6.2	10,173	5.0	n.a	
Limited English proficiency with accommodations	6,867	3.5	7,272	3.5	n.a	
Limited English proficiency without accommodations	2,974	1.5	2,017	1.0	n.a	
Individualized education plan with or without accommodations	n.a		n.a		6,757	5.8
Limited English proficiency with or without accommodations	n.a		n.a		2,345	2.0
Transferred schools during the school year	12,744	6.5	13,074	6.4	4,962	4.3

Source: OSDE administrative data.

Notes: All percentages are based on the counts in the top row.

Because relatively few students had test accommodations on EOI post-tests, students with and without accommodations were pooled when accounting for individual education plans and limited English proficiency for EOI subjects.

For all student characteristics in this table, less than 1 percent of students have missing data.

n.a. = not applicable

C. Dosage for teacher-student links

Some students were taught by more than one teacher, either because they moved between schools or were taught by multiple teachers in the same school. We refer to the fraction of the time a student was taught by a given teacher for a subject as the dosage. In this section, we describe how we calculated dosage for the value-added model.

1. Roster verification

In the 2012–2013 school year, OSDE implemented a pilot roster verification program in selected schools, which was expanded to additional schools in the 2013–2014 school year.[8] Roster verification is a process by which records of teachers' monthly shares of instruction for each student and course are submitted and either verified or corrected by teachers and school administrators. For example, consider a student who spends 2.5 days per week in teacher A's classroom learning math and 2.5 days per week in teacher B's classroom learning math. This student would be recorded as having spent 50 percent of math instructional time with teacher A for that month. Likewise, the same student would also be recorded as having spent 50 percent of math instructional time with teacher B for that month. In recording their share of instructional time with a student, teachers rounded to the nearest quarter. Thus, 0, 25, 50, 75, and 100 percent were the possible responses. After these shares were reported, they were verified or corrected by those teachers and school administrators. The roster verification process differed slightly in the Tulsa public schools, where teachers rounded to the nearest 10 percent. Roster verification was not implemented statewide and was not always fully implemented in the pilot schools.

2. Dosage

Teacher dosage measures the proportion of instructional time a teacher spent with a student during the school year. We excluded instructional time after the post-tests from the calculation of dosage. For example, because the spring testing window spans April to early May, we excluded instructional time in May or later from the calculation of dosage. Similarly, we excluded instructional time in January or later for students who took an EOI post-test during the winter testing window in December.[9] To calculate teacher dosage for a teacher-student link, we used a three-step process: (1) we determined the amount of instructional time in each month of the school year, (2) we summed the monthly dosages, and (3) we divided by the number of months to obtain dosage as a percentage of the school year through April (or December for post-tests taken in the winter testing window).

So that solo-taught and co-taught students contribute equally to teachers' value-added estimates, we assigned each teacher full credit for the shared students when two or more teachers claimed the same students at 100 percent during the same term. In doing so, we summed monthly dosages for a teacher-student combination when the student was linked to the teacher in multiple courses in the same subject. Thus a teacher-student link could have a dosage that exceeds 100 percent. For example, a teacher who claimed the same student in roster confirmation for two full-year courses, and assigned 100 percent instructional time in every month for both courses would have a combined dosage of 200 percent for that student.

---

[8] The full rollout of roster verification for Oklahoma is planned for the 2014–2015 school year.

[9] We calculate dosage the same way for all students who take the test during the same testing window because we know only the season of the testing window and do not know the exact date individual students were tested. In addition to the winter and spring testing windows, some Oklahoma students take the EOI assessments on a trimester schedule. These students take the test in late January or early February. We calculate dosage for students taking an EOI test during this window the same as for students taking a test during the winter window because we cannot distinguish between tests taken during these two testing windows.

Although we obtained monthly dosage from the roster-verified records when possible, roster verification was not implemented statewide. For teachers without verified roster data, we assumed that the monthly dosage for a teacher-student link was equal to the proportion of instructional days the student was officially enrolled in the school based on administrative data from OSDE. These data contained dates of school withdrawal and admission.

For dosage based on both roster-verified and administrative teacher-student links, we assume that learning accumulated at a constant rate and, therefore, treat days spent at one school as interchangeable with days spent at another. For example, if a student split time equally between two schools, we set the dosage of each school to 50 percent, regardless of which school the student attended first.

3. Catch-all teachers

Some students appear to not be linked to a math or reading/ELA teacher because roster verification was not implemented statewide and because of limitations in the administrative data provided by OSDE that linked teachers to students. For example, some students with post-test scores in the analysis file were not linked to a teacher in the post-test's subject. All such students were assigned a placeholder teacher for each subject in which they had no roster record linking them to a teacher. We created a so-called catch-all teacher in OCCT subjects for each school-grade combination that had unlinked post-tested students. We did the same in EOI subjects for each school that had unlinked post-tested students. Teacher dosages were assigned to catch-all teachers in the same manner as teachers with unverified roster records. Across all grades and subjects, 5.6 percent of teacher-student links are to catch-all teachers of unassigned students.

This page has been left blank for double-sided copying.

III. ESTIMATING TEACHER VALUE ADDED

In this chapter, we describe the technical details of the teacher value-added model. Our approach is to obtain initial estimates of teacher effectiveness from linear regressions. We then obtain final value-added results by applying two adjustments to the initial estimates to account for differences across grades and different numbers of students per teacher. We organize the discussion into four topics: (1) the estimation equations, (2) how we address measurement error in the pre-tests in the regressions, (3) how we generalize estimates to be comparable across grades, and (4) how we account for the number of students per teacher.

A. Estimation equations

We developed two linear regression approaches to estimate school and teacher effectiveness measures, based on whether the post-test score was from the OCCT or EOI. For the OCCT approach, we estimated regression models separately for each grade and subject combination. This approach covered grades 4 through 8 and the subjects reading and math. For the EOI approach, we estimated the regression models separately for each subject only. The subjects were algebra I, geometry, algebra II, and English III. For these EOI subjects, we pooled the regression models by subject across grades because the assessment for a given subject is the same in any grade that it is taken. For EOI subjects, the grades were 8 and 9 for algebra I, 9 through 11 for geometry, 9 through 12 for algebra II, and 11 for English III. We excluded grades other than these for EOI subjects because too few students had post-tests.

For both the OCCT and EOI approaches, the post-test score depends on pre-test scores, student background characteristics, the student's teacher, and unmeasured factors. For a given teacher *t* and student *i* in grade *g*, the regression equation for the OCCT approach is:

$$(1)\quad Y_{tig} = \lambda_{1Mg} M_{i(g-1)} + \lambda_{1Rg} R_{i(g-1)} + \lambda_{1Sg} S_{i(g-1)} + \boldsymbol{\beta}'_{1g}\mathbf{X}_i + \boldsymbol{\delta}'_{1g}\mathbf{T}_{1tig} + \varepsilon_{1tig}\ ,$$

and the regression equation for the EOI approach is:

$$(2)\quad Y_{tig} = \boldsymbol{\lambda}'_2\mathbf{P}_{i(g-1)} + \boldsymbol{\gamma}'\mathbf{C}_{i(g-1)} + \boldsymbol{\kappa}'\mathbf{G}_g + \boldsymbol{\tau}'\mathbf{A}_i + \boldsymbol{\beta}'_2\mathbf{X}_i + \boldsymbol{\delta}'_2\mathbf{T}_{2ti} + \varepsilon_{2tig}\ ,$$

In both equations, Y_{tig} is the post-test score. The regressions are run separately by post-test subject, so we drop a subject subscript for ease of notation. In the OCCT equation, $M_{i(g-1)}$ is the math pre-test for student *i* from the previous grade *g-1*, and $R_{i(g-1)}$ is the reading pre-test from the previous grade. For students in grade 6, we also include $S_{i(g-1)}$, the science pre-test taken in the previous grade.

In the EOI equation, the vector $\mathbf{P}_{i(g-1)}$ denotes variables for pre-test scores in each of the included subjects in the previous grade. Unlike the OCCT approach, in the EOI approach not all students with a post-test in a given subject have scores from the same pre-tests. For example, the regression for the geometry post-test included five pre-test variables—grade 8 math, algebra I, algebra II, English II, and biology. Because all students included in the EOI models were required to have a pre-test score on a same-content assessment, students included in the regression for the geometry post-test have a pre-test score in one of grade 8 math, algebra I, or algebra II. Some students in the geometry regression additionally have pre-test scores in English II and/or biology, but these were not required for all students. To account for different pre-test subjects, the EOI equation includes binary variables in $\mathbf{C}_{i(g-1)}$ that indicate whether a student had a pre-test in each subject. The vector $\mathbf{C}_{i(g-1)}$ also includes variables that indicate whether the pre-test was the second time a student had taken the test in the subject. For example, a student who took algebra I in 2011–2012, algebra I again in 2012–2013 (the pre-test), and geometry in 2013–2014 (the post-test) would be indicated in $\mathbf{C}_{i(g-1)}$ as having taken algebra I as a pre-test and as having taken algebra I a second time. We excluded students who were retaking a post-test they took the year before.

The pre-test scores in both equations capture prior inputs into student achievement; we estimated the associated coefficients—λ_{1Mg}, λ_{1Rg}, λ_{1Sg}—and the vector $\boldsymbol{\lambda}_2$, using a procedure that corrects for measurement error in these pre-test scores. The subscripts 1 and 2 distinguish the OCCT model coefficients from the EOI model coefficients.

The vector $\mathbf{X}_i$ denotes the control variables for student background characteristics. For the OCCT approach, we allowed these coefficients to vary by grade, represented in equation 1 by the g subscript on the vector $\boldsymbol{\beta}_{1g}$. Because there are some grade and subject combinations for EOI post-test scores with very few students, we did not allow the coefficients to vary by grade for the EOI approach. Doing so leads to more precise coefficient estimates in $\boldsymbol{\beta}_2$ and could lead to more precise value-added results.

The vectors $\mathbf{T}_{1tig}$ and $\mathbf{T}_{2ti}$ consist of binary variables for each teacher. Because the OCCT approach is estimated separately by grade, a teacher who taught multiple grades had variables in each grade regression model. For example, a teacher who taught math in grades 4 and 5 had one variable in $\mathbf{T}_{1ti4}$ for the grade 4 regression and one in $\mathbf{T}_{1ti5}$ for the grade 5 regression. The EOI approach is not grade-specific, so a teacher who taught multiple grades had only one variable in $\mathbf{T}_{2ti}$. Each teacher-student observation has one nonzero element in $\mathbf{T}_{1tig}$ and/or $\mathbf{T}_{2ti}$. The coefficient vectors $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ contain the initial estimates of teacher effectiveness. Rather than dropping one element of $\mathbf{T}_{1tig}$ or $\mathbf{T}_{2\mathrm{ti}}$ from the regression, we estimated the regression models without constant terms. The vectors $\mathbf{T}_{1tig}$ and $\mathbf{T}_{2ti}$ also include binary variables for the catch-all teachers of unassigned students. We also mean-centered the control variables so that each element of $\boldsymbol{\delta}_{1g}$ and $\boldsymbol{\delta}_2$ represents a teacher-specific intercept term for a student with average characteristics.

We also accounted for factors beyond a teacher's control that might drive cross-grade differences in value added. For the EOI approach, we included binary variables for each grade in the vector $\mathbf{G}_g$. We excluded the indicator for the highest included grade for each subject. These variables account for the possibility that a student's achievement in a subject could be related to their grade level. This approach avoids penalizing teachers for teaching in grades with lower-ability students. Because we estimated the OCCT model separately by grade, the approach does not include grade variables. However, we adjusted for possible differences across grades in the average and variability of value-added estimates, using the approach described in Section III.C.

Finally, to account for any differences in achievement that may be related to the timing of EOI tests, we also included the vector $\mathbf{A}_i$ with indicators for the season the student took the post-test (winter, spring, or summer).

Table III.1 shows the coefficient estimates and standard errors of the control variables in the model by subject and grade span. The top panel shows the average association between the pre-tests and achievement on the post-tests (measured in points on the test), accounting for all other variables in the regression. The bottom panel shows the association between each student characteristic and post-test scores.

To account for team teaching, we used the Full Roster Method (Hock and Isenberg 2012). In this approach, each student contributed one observation to the model for each teacher to whom he or she was linked. Thus, the unit of observation in the analysis file is a teacher-student combination. This method of accounting for team teaching is based on the assumption that teachers who contribute equally to student achievement within each team receive equal credit.

Because some students contributed multiple observations, we estimated the coefficients by using weighted least squares (WLS) rather than ordinary least squared (OLS). We weighted each record based on the dosage associated with the teacher-student combination. In this model, the error terms are correlated, because individual students have multiple records, and heteroskedastic, due to differences across students in how well the model can predict post-test scores based on the background characteristics included in the regressions. Therefore, we used a cluster-robust sandwich variance estimator (Liang and Zeger 1986; Arellano 1987) to produce consistent standard errors in the presence of heteroskedasticity and correlation in the regression error term.

The regression models yield initial estimates of teacher effectiveness for each grade and subject for the OCCT approach and for each subject in the EOI approach. For the OCCT approach, we included teachers in the regression model only if they had at least five students in that grade and subject combination. For the EOI approach, we calculated initial estimates for teachers with at least five students across all eligible grades. We used the initial estimates in the subsequent step to generalize across grades, then applied empirical Bayes (EB) shrinkage to the generalized estimates. As a final step, we removed any teachers with fewer than 10 students across all eligible grades.

How to Interpret Table III.1

Table III.1 displays the regression coefficients from the value-added model. These coefficients describe the relationships between the characteristics of Oklahoma students and achievement on the post-test. The coefficients give the amount of the increase (or decrease if the coefficient is negative) in the typical achievement of students when a characteristic increases by one unit. For example, the coefficient of 0.66 in the first row of the second column of the table indicates that an increase by one OCCT point on a student's pre-test score is associated with a typical increase of 0.66 points on the OCCT math post-test. Similarly, the coefficient on the fraction of the prior year a student attended school indicates that the typical score for a student who attended 100 percent of the prior year is 42 points higher than the typical score for a student who instead had attended for none of the prior year. More than 99.5 percent of students attended 80 percent or more of the year before, so the usual contribution of previous attendance to students' typical scores is much smaller than this change of 42 points might suggest; the change in typical scores associated with a change in attendance from 80 to 100 percent is 8.4 OCCT points.

For characteristics that are yes/no indicators, the coefficient gives the increase in the typical score for a student who has that characteristic relative to an otherwise similar student who does not. For example, the typical score for students in grades 4 through 8 math who transferred between schools during the school year is 7.16 points lower than for students who did not transfer. The other binary indicators are in groups of related indicators. For example, the coefficients on the two indicators of student poverty status measure the difference in the typical score of a student with that status (for example, students eligible for reduced-price lunch) relative to a student who is eligible for free lunch.

Each regression coefficient describes a relationship after accounting for all other characteristics included in the model. Accordingly, the coefficient on a characteristic gives the change in typical achievement when the characteristic is changed from no to yes or increased by one point, assuming that all of the students' other characteristics remain the same. Consequently, coefficients may not reflect the relationship we would observe had the other characteristics not been accounted for in the value-added model. This feature of multiple regression coefficients can produce counterintuitive relationships between characteristics and achievement if the contributions of one characteristic are accounted for largely by another characteristic in the model. For example, coefficients on limited English proficiency status would likely be consistently negative and greater in magnitude if the model did not also account for students' pre-test scores, because students with limited English proficiency tend to have lower pre-test scores.

The magnitude of the coefficients can be compared to the typical range of student achievement on an OCCT or EOI test. The standard deviation of student achievement on the grade 4 math post-test was 90.9 OCCT points, indicating that about two-thirds of students scored within 90.9 points above or below the average score on the assessment. The standard deviations for other grades ranged from 71.9 to 90.3 points in grades 4 through 8 math, from 68.2 to 54.8 points in grades 4 through 8 reading, and from 50.2 to 70.1 points in the four EOI subjects.

The number in parentheses below each coefficient is the *standard error of the coefficient*—a measure of precision. A more precise coefficient indicates with more certainty that a coefficient reflects the actual relationship between the characteristic and achievement. Coefficients with smaller standard errors are more precise. The coefficients on the pre-tests are more precise than those on individual background characteristics. Roughly, a coefficient that is at least twice as large as its standard error is said to be statistically significant, meaning that it is likely that the direction of the relationship—whether positive or negative—reflects the actual relationship between the characteristic and achievement and is unlikely to be produced by chance.

Table III.1. Coefficients on covariates in the value-added models, by post-test subject

Variable	OCCT Math	OCCT Reading	Algebra I	Geometry	Algebra II	English III
Pre-test scores (average coefficients)						
Math	0.66 (0.01)	0.15 (0.01)	0.40 (0.01)	0.76 (0.01)	0.59 (0.01)	0.06 (0.01)
Reading/ELA	0.15 (0.01)	0.65 (0.01)	0.06 (0.01)	0.05 (0.01)	0.04 (0.01)	0.62 (0.01)
Science	0.15 (0.01)	0.31 (0.01)	0.11 (0.01)	0.15 (0.01)	0.10 (0.01)	0.07 (0.01)
Individual student background characteristics (average coefficients)						
Ineligible for free or reduced-price lunch	4.35 (0.64)	5.26 (0.62)	1.06 (0.45)	1.87 (0.56)	-0.17 (0.70)	1.75 (0.52)
Eligible for reduced-price lunch	1.52 (0.95)	2.22 (0.90)	0.22 (0.68)	1.67 (0.89)	0.57 (1.10)	0.24 (0.85)
Female	1.39 (0.52)	5.85 (0.50)	5.25 (0.38)	-3.00 (0.47)	4.95 (0.56)	1.80 (0.43)
African American	-3.06 (1.22)	-1.20 (1.18)	-0.16 (0.94)	-5.30 (1.22)	1.84 (1.44)	-1.76 (1.14)
Hispanic	0.83 (0.90)	-0.87 (0.85)	2.02 (0.64)	-1.92 (0.84)	2.65 (1.03)	0.61 (0.75)
American Indian	-1.24 (0.99)	0.25 (0.96)	-1.64 (0.74)	-0.62 (0.96)	-0.60 (1.12)	-1.26 (0.91)
Asian/Pacific Islander	12.66 (1.87)	3.78 (1.70)	9.26 (1.37)	5.20 (1.75)	9.21 (1.83)	1.62 (1.56)
Caucasian/other	0.19 (1.04)	1.62 (1.01)	-2.07 (0.80)	1.47 (1.02)	1.11 (1.21)	-0.03 (0.95)
Individual education plan with accommodations	-4.02 (1.38)	-3.81 (1.21)	n.a.	n.a.	n.a.	n.a.
Individual education plan without accommodations	-5.55 (1.39)	-11.80 (1.43)	n.a.	n.a.	n.a.	n.a.
Limited English proficiency with accommodations	-0.73 (1.81)	-4.75 (1.60)	n.a.	n.a.	n.a.	n.a.
Limited English proficiency without accommodations	-2.08 (3.10)	-7.60 (3.51)	n.a.	n.a.	n.a.	n.a.
Individual education plan with or without accommodations	n.a.	n.a.	-2.17 (0.96)	-11.42 (1.41)	-8.96 (1.69)	-4.23 (1.18)
Limited English proficiency with or without accommodations	n.a.	n.a.	2.90 (1.50)	-0.73 (2.27)	6.97 (3.17)	2.22 (2.59)
Transferred schools during the school year	-7.16 (1.26)	-4.79 (1.18)	-6.17 (0.99)	-7.95 (1.41)	-14.27 (2.31)	-2.98 (1.21)
Fraction of the prior year student attended school	0.42 (0.10)	-0.21 (0.10)	0.35 (0.07)	0.40 (0.09)	0.67 (0.11)	0.02 (0.07)

Source: Mathematica calculations based on OSDE administrative data.

Notes: Standard errors are in parentheses.

For OCCT post-tests, the reported coefficient estimates represent weighted averages of the coefficients estimated separately for each grade, where the weights are the number of student equivalents in the grade. Additionally, for EOI post-tests, the reported coefficient estimates of pre-test scores represent weighted averages of the coefficients estimated separately for each pre-test subject in the content area, where the weights are the number of student equivalents with a pre-test score in a given subject. The associated standard errors similarly represent weighted averages across grades or subjects. These numbers are presented for descriptive purposes only and should not be used to conduct rigorous statistical tests.

The math pre-test score for students in the OCCT math value-added model is the OCCT math assessment from the previous grade. The math pre-test score for students in the EOI models is another math OCCT or EOI assessment from the previous grade. Similarly, reading/ELA pre-test scores are OCCT reading or English II scores from the previous grade. OCCT science pre-test scores are included for students in grades 6 and 9, and EOI biology pre-test scores are included for students who took the test in the previous grade.

Coefficients on the poverty status variables are relative to students eligible for free lunch—the excluded category. Students can be indicated in multiple race/ethnicity categories, so the coefficients on the variables are relative to students who are not in the specific race/ethnicity category.

n.a. = not applicable.

B. Measurement error in the pre-tests

We corrected for measurement error in the pre-tests by using data on test reliability. As a measure of true student ability, student achievement tests contain measurement error. This error causes standard models to produce biased estimates of teacher effectiveness. To address this issue, we implemented an errors-in-variables correction (Buonaccorsi 2010). Using information about the reliability of the OCCT and EOI tests, available from the test publisher, the correction nets out the known amount of measurement error (CTB/McGraw-Hill 2013a, 2013b).

Correcting for measurement error required two regression steps because of computational limitations with the measurement-error correction method related to producing measures of precision. In the first step, we applied the errors-in-variables correction to account for measurement error. The second regression step was necessary to calculate standard errors on teachers' estimates.

In the first regression step, we used the errors-in-variables approach to get the initial estimates of value added. We estimated the regression equations 1 and 2 with the correction to obtain unbiased estimates of the coefficients on the pre-test scores. We based the correction on the published reliabilities for each OCCT and EOI test. For the OCCT model, we used grade- and subject-specific reliability data. For the EOI regression, we used subject-specific reliability data. We then used the coefficients to calculate an *adjusted* post-test score that nets out the contribution of the pre-test scores. The adjusted post-test score for the OCCT approach is given by:

$$(3) \quad A_{1tig} \equiv Y_{tig} - \lambda_{Mg} M_{i(g-1)} - \lambda_{Rg} R_{i(g-1)} - \lambda_{Sg} S_{i(g-1)}$$

The adjusted post-test scores for the EOI approach are given by:

$$(4) \quad A_{2tig} \equiv Y_{tig} - \boldsymbol{\lambda}_2' \mathbf{P}_{i(g-1)}$$

The vectors A_{1tig} and A_{2tig} represent the post-test scores, net of the estimated contribution of the student's pre-test scores. We calculated an adjusted post-test score for each OCCT grade and subject and for each EOI subject.

We used these adjusted post-test scores in a second regression step to obtain standard errors that are consistent in the presence of both heteroskedasticity and clustering at the student level, because the regression includes multiple observations for the same student. This second-stage regression is necessary because it is not computationally possible to simultaneously account for correlation in the error term ε across multiple observations and apply the numerical formula for the errors-in-variables correction. Thus, for each OCCT grade and subject we estimated the final regression in equation 5:

$$(5) \quad A_{1tig} = \boldsymbol{\beta}_{1g}' \mathbf{X}_i + \boldsymbol{\delta}_1' \mathbf{T}_{tig} + \varepsilon_{1tig} .$$

and for each EOI subjects, we estimated the final regression in equation 6:

$$(6) \quad A_{2tig} = \boldsymbol{\gamma}'\mathbf{C}_{i(g-1)} + \boldsymbol{\kappa}'\mathbf{G}_g + \boldsymbol{\beta}_2'\mathbf{X}_i + \boldsymbol{\delta}_2'\mathbf{T}_{2ti} + \varepsilon_{2tig}\ .$$

The coefficients appear in equations 5 and 6 as they did in equations 1 and 2 because the regressions produce identical coefficient estimates; equations 5 and 6 apply a correction only to the standard errors.

This two-step method likely underestimates the standard error of the estimated $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ because the adjusted gains in equations 3 and 4 rely on the estimated values of the pre-test coefficients λ_{Mg}, λ_{Rg}, λ_{Sg}, and the vector $\boldsymbol{\lambda}_{(g-1)}$. Treating these coefficients as fixed rather than as estimates does not fully account for variability in post-test scores related to pre-test scores. Nonetheless, with the large within-grade and within-subject sample sizes, the pre-test coefficients were precisely estimated, likely leading to a negligible difference between the standard errors we obtained for the teacher value-added results and those we would have obtained had we not needed to treat the pre-test coefficients as fixed.

Underestimated standard errors could result in insufficient shrinkage of some teachers' value-added estimates, which we discuss in Section III.D. When using value-added point estimates for teacher evaluations, the key concern is not whether the standard errors of the estimates are universally underestimated, but whether the standard errors for some teachers are disproportionately underestimated, which can lead to some teacher estimates shrinking too little relative to other teacher estimates in the final step. Thus, there is a tradeoff in the design of the model between insufficient shrinkage for some teachers and accounting for measurement error. This approach emphasizes accuracy and face validity of teachers' value-added estimates over any consequences of underestimated standard errors for the shrinkage procedure.

C. Generalizing estimates to make them comparable across grades

Both the average and variability of the initial value-added estimates may differ across grade levels, leading to a potential problem when comparing teachers assigned to different grades. The main concern is that factors beyond teachers' control might drive cross-grade discrepancies in the distribution of value-added estimates. For example, the standard deviation of adjusted post-test scores might vary across grades because of differences in the alignment of tests or in knowledge retention from one year to the next. However, in TLE, all teachers will be compared within a subject, regardless of any grade-specific factors that might affect the distribution of gains in student performance between years.

Because of differences in our approach to estimating value added based on the OCCT and EOI tests, our method to address differences across grades also varied. For the OCCT approach, we transformed the grade-specific value-added estimates to be comparable across grades and then combined these transformed estimates for teachers of multiple grades. For the EOI approach, we addressed differences across grades by accounting for grade in the regression model.

1. Grade-level adjustments in the OCCT approach

Transforming estimates into generalized OCCT points. For value-added results based on the OCCT tests, we translated teachers' grade-level estimates so that each set of estimates is expressed in a common metric of generalized OCCT points. Aside from putting value-added estimates for teachers on a common scale, this approach leads to distributions of teacher estimates that are more equal across grades. Doing so avoids penalizing or rewarding teachers simply for teaching in a grade with atypical test properties. However, the approach does not reflect a priori knowledge that the true distribution of teacher effectiveness is similar across grades. Rather, without a way to distinguish cross-grade differences in teacher effectiveness from cross-grade differences in testing conditions, the test instrument itself, or student cohorts, the approach reflects an implicit assumption that the distribution of true teacher effectiveness is the same across grades.

We standardized the initial estimates of teacher effectiveness from the OCCT regressions so that the mean and standard deviation of the distribution of teacher estimates is the same across grades. First, we subtracted from each initial estimate the average of all estimates within the same grade. We then divided the result by an estimate of the standard deviation within the same grade. To reduce the influence of imprecise estimates obtained from teacher-grade combinations with few students, we calculated the average using weights based on the number of students taught by each teacher. Our method of calculating the standard deviation of teacher effects also assigns less weight to imprecise individual estimates. Finally, we multiplied by the square root of the teacher-weighted average of the grade-specific variances, obtaining a common measure of effectiveness on the generalized OCCT-point scale.

Formally, the value-added result expressed in generalized OCCT points is the following:

$$(7)\quad \hat{\eta}_{tg} = \frac{\left(\hat{\delta}_{tg} - \overline{\hat{\delta}_g}\right)}{\hat{\sigma}_g} \times \sqrt{\left(\frac{1}{K}\sum_{h=4}^{8} K_h \hat{\sigma}_h^2\right)},$$

where $\hat{\delta}_{tg}$ is the grade-*g* estimate for teacher *t*, $\overline{\hat{\delta}_g}$ is the weighted average estimate for all teachers in grade *g*, $\hat{\sigma}_g$ is the estimate of the standard deviation of teacher effectiveness in grade *g*, K_h is the number of teachers with students in grade *h*, and *K* is the total number of teachers.

In equation 7, we used an adjusted standard deviation that removes estimation error to reflect the dispersion of underlying teacher effectiveness. The unadjusted standard deviation of the value-added estimates will tend to overstate the true variability of teacher effectiveness; because the scores are regression estimates, rather than known quantities, the standard deviation will partly reflect estimation error. Using the unadjusted standard deviations to scale estimates could lead to over- or underweighting one or more grades when the extent of estimation error differs across grades. This is because doing so would result in estimates with the same amount of total dispersion—the true variability of teacher effectiveness and the estimation error combined—in each grade, but the amount of true variability in each grade would not be equal. Instead, we scaled the estimates using the adjusted standard deviation, spreading out the distribution of effectiveness in grades with relatively imprecise estimates so that estimates of teacher effectiveness in each grade have the same true standard deviation.[10]

We calculated the error-adjusted variance of teacher value-added results separately for each grade as the difference between the weighted variance of the grade-*g* teacher estimates and the weighted average of the squared standard errors of the estimates. The error-adjusted standard deviation $\hat{\sigma}_g$ is the square root of this difference. We chose the weights based on the EB approach outlined by Morris (1983). In this approach, the observed variability of the value-added estimates is adjusted downward according to the extent of estimation error.

Table III.2 shows the adjusted standard deviation of the initial estimates of teacher effectiveness derived from the value-added regression, as well as the weighted average across all grades produced by equation 7. A higher standard deviation for a grade-year combination indicates more dispersion in underlying teacher effectiveness before the estimates were transformed into generalized OCCT points. The standard deviation of value added ranged from 21.3 to 22.1 OCCT points in math and from 11.2 to 12.3 points in reading. By comparison, the range of the standard deviations of student-level achievement across grades was 71.9 to 90.9 OCCT points in math and 68.2 to 84.8 points in reading. Because we estimated value-added results for EOI subjects pooling all grades, we report only the combined standard deviations. These ranged from 5.7 to 24.3 EOI points. By comparison, the range of the standard deviations of student-level achievement across grades was 50.2 to 70.1 EOI points.

Table III.2. Adjusted standard deviations of value added, by subject and grade

	Grade						
Model	4	5	6	7	8	Weighted average	Student-level achievement
OCCT math	22.1	21.7	21.8	21.4	21.3	22.0	71.9 - 90.9
OCCT reading	12.1	11.8	12.3	11.2	11.5	12.0	68.2 - 84.8
Algebra I						15.0	50.2
Geometry						24.3	70.1
Algebra II						14.2	62.9
English III						5.7	51.1

Source: Mathematica calculations based on OSDE administrative data.

Notes: Teachers are included in the calculation of the standard deviation for each grade that they teach. The average standard deviation is weighted by the number of teachers in each grade.

Combining OCCT estimates for teachers of multiple grades. To combine grade-level estimates from OCCT models into a single (unshrunk) value-added result, denoted as $\hat{\eta}_t$, for a teacher with students in multiple grades, we used a weighted average of the grade-specific estimates (expressed in generalized OCCT points). We set the weight for grade *g* equal to the proportion of students of teacher *t* in grade *g*. Because combining teacher effects across grades

[10] For teachers in grades with imprecise estimates, the shrinkage procedure, described in Section III.D, counteracts the tendency for these teachers to receive final estimates in the extremes of the distribution.

might cause the overall average to be nonzero, we recentered the estimates on zero before proceeding to the next step.

We computed the variance of each teacher's combined effect as a weighted average of the grade-specific squared standard errors of the teacher's estimates. We set the weight for grade *g* equal to the squared proportion of students of teacher *t* in grade *g*. For simplicity, we assumed that the covariance across grades is zero. In addition, we did not account for uncertainty arising because $\overline{\hat{\delta}_g}$ and $\hat{\sigma}_g$ in equation 7 are estimates of underlying parameters rather than known constants. Both decisions imply that the standard errors will be underestimated slightly.

2. Grade-level adjustments in the EOI approach

Unlike the grade-by-grade OCCT approach, we pooled grades for the EOI estimation equation and estimated a single initial value-added estimate for each teacher, rather than one for each teacher-grade combination. We also used a different method to account for differences across grades because of this difference in approach. To account for differences across grades in EOI models, we included binary variables for each grade (excluding the highest grade) in the regressions. For example, students taking geometry in grade 9 or 10 might be higher-ability students on average compared to students taking geometry in grade 11, even after accounting for the other variables in the model. The coefficients on the grade 9 and grade 10 indicators would give the increase in the typical achievement for students in these grades relative to a student taking geometry in grade 11. This approach avoids penalizing teachers for teaching in grades with lower-ability students.

One potential concern with this approach is that the coefficients could also measure differences in the effectiveness of teachers in different tracks, which could lead to bias in the value-added results related to teachers' grade assignments. This is unlikely, because the coefficients on the grade indicators are based on comparing the achievement of students whose teachers teach the same subject to students in multiple grades, instead of comparing achievement across teachers. We achieved this by simultaneously including $\mathbf{T}_{2ti}$ in the regression equation so that all relationships between the variables included in the regression and achievement—including the grade indicators—were based on within-teacher variation in student achievement, rather than on variation from students in different teachers' classrooms.

D. Accounting for different numbers of students

To reduce the risk that teachers, particularly those with relatively few students in their grade, will receive a very high or very low effectiveness measure by chance, we applied the EB shrinkage procedure (Herrmann et al. 2013). Using the EB procedure outlined in Morris (1983), we computed a weighted average of an estimate for the average teacher and the initial estimate based on each teacher's own students. For teachers with relatively imprecise initial estimates based on their own students, the EB method effectively produces an estimate based more on the average teacher. For teachers with more precise initial estimates based on their own students, the EB method puts less weight on the value for the average teacher and more weight on the value obtained from the teacher's own students.

The EB estimate for a teacher is approximately equal to a precision-weighted average of the teacher's estimated effect (after generalizing across grades) and the overall mean of all estimated teacher effects. [11] Following the standardization procedure, the overall mean is zero, with better-than-average teachers having positive scores and worse-than-average teachers having negative scores. We therefore arrived at the following:

$$(8)\quad \hat{\eta}_t^{EB} \approx \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\sigma}_t^2} \right) \hat{\eta}_t ,$$

where $\hat{\eta}_t^{EB}$ is the EB estimate for teacher *t*, $\hat{\eta}_t$ is the initial estimate of effectiveness for teacher *t* based on the regression model (after combining OCCT estimates across grades), $\hat{\sigma}_t$ is the standard error of the estimate of teacher *t*, and $\hat{\sigma}$ is an estimate of the standard deviation of teacher effects (purged of sampling error), which is constant for all teachers. Equation 8 has no explicit term for the weight on the overall mean because this mean is zero. The term $[\hat{\sigma}^2/(\hat{\sigma}^2 + \hat{\sigma}_t^2)]$ must be less than one. Thus, the EB estimate always has a smaller absolute value than the initial estimate—that is, the EB estimate "shrinks" from the initial estimate. The greater the precision of the initial estimate—that is, the smaller $\hat{\sigma}_t^2$ is—the closer $[\hat{\sigma}^2/(\hat{\sigma}^2+\hat{\sigma}_t^2)]$ is to one and the smaller the shrinkage in $\hat{\eta}_t$. Conversely, the larger the variance of the initial estimate, the greater the shrinkage in $\hat{\eta}_t$. By applying a greater degree of shrinkage to less precisely estimated teacher measures, the procedure reduces the likelihood that the estimate of effectiveness for a teacher falls at either extreme of the distribution by chance. We calculated the standard error for each $\hat{\eta}_t^{EB}$ using the formulas provided by Morris (1983). As a final step, we removed any teachers with fewer than 10 students and recentered the EB estimates on zero.

[11] In Morris (1983), the EB estimate does not exactly equal the precision-weighted average of two values due to a correction for bias. This adjustment decreases the weight on the estimated effect by a factor of (K–3)/(K–1), where K is the number of teachers. For ease of exposition, we have omitted this correction from the description given here.

This page has been left blank for double-sided copying.

IV. SCHOOL AND SUBGROUP VALUE-ADDED RESULTS FOR TLE

Educators will receive three types of results derived from the value-added results for teachers: (1) TLE component scores, (2) school value-added results, and (3) value-added results for student subgroups. In this chapter, we describe how we calculate each of these statistics from the value-added results.

A. Using the value-added results in TLE

We provided OSDE with the original generalized OCCT or EOI point value-added results, the average typical and actual scores, and value-added results that had been converted to a scale from 1.0 to 5.0 for each teacher in the model. TLE requires that the value-added results take on one of 41 values from 1.0 to 5.0 when they are used to calculate composite TLE ratings. The Oklahoma State Board of Education approved the TLE Commission's recommended method of converting the value-added results to this scale. In this system, the average Oklahoma teacher received a score of 3.0; teachers whose results exceeded the average by two standard deviations received a score of 5.0; and those whose results fell below the average by two standard deviations received a score of 1.0. Teachers who were eligible for value-added results in multiple subjects received scores in each subject. We then assigned these teachers a combined component score that was a weighted average of their subject-specific scores. The weight given to each subject was the number of student equivalents.

B. School value-added results

Each school's value-added result reflects the combined contributions of teachers at that school. For each subject, we calculated a school's value added as a weighted average of the school's initial teacher value-added results.[12] The weight given to each teacher was the number of student equivalents. Rather than giving equal weight to each teacher in the calculation of school value added, this approach gives equal weight to students who have the same dosage in the value-added model. The averages for schools also include value-added results for the catch-all teachers of unassigned students. We then applied the same two post-regression adjustments to the school estimates that we did for teachers; we standardized OCCT estimates across grades, and applied shrinkage to the combined estimates. We removed estimates in a subject for schools with fewer than 10 students in the subject. As a final step, we calculated TLE component scores for schools across all subjects using the same procedure we used for teachers.

This method for calculating school value-added results provides two benefits that other commonly used methods do not. First, this method directly measures the contributions of teachers in the school, making this approach more transparent than some alternatives. Second, the primary alternative approach to estimating school value added—estimating a regression similar to that for the teacher model, but replacing teacher indicators with school indicators—

[12] The initial estimates used in to calculate the school's average differ slightly from those produced by the models described in Chapter III. To measure distinct contributions to the achievement of students in different schools of teachers who taught at multiple schools, the regressions were modified to include indicators for each teacher-school combination instead of only for each teacher. Additionally, only students linked to teacher-school combinations that met the five-student minimum were included in the regressions. We included catch-all teachers of unassigned students in the school value-added model just as we did for the teacher value-added model.

might not distinguish achievement due to the characteristics of students in the school from achievement due to the effectiveness of teachers in the school. This is because omitting the teacher indicators can lead to estimated relationships between student background characteristics and achievement that do not fully account for how students are assigned to teachers within schools, thereby attributing achievement caused by factors outside of teachers' control to schools. Our approach addresses this concern by averaging direct estimates of teachers' contributions.

C. Value-added results for student subgroups

We used the students' actual post-test scores and estimated typical scores to calculate the value-added results for student subgroups. The teacher's (or school's) value-added result for a subgroup is the difference between the average actual post-test scores and the average typical scores, where the averages are calculated based only on scores for students in the subgroup. The typical scores reflect the adjustments that we made to the initial value-added estimates described in Chapter III, including shrinkage and standardization across grades. We mean-centered the subgroup value-added results by combinations of subgroups and subjects so that a positive subgroup result reflects above-average contributions to the achievement of students in the subgroup compared to other Oklahoma teachers or schools. As a final step, we removed subgroup value-added results for teachers or schools with fewer than seven students in the subgroup.[13]

For the reporting of subgroup results for teachers and schools, we transformed the results to take on values of above average, average, or below average. We defined above average as value-added results that are greater than or equal to one standard deviation above the average teacher value-added result in the subject-subgroup combination. Similarly, we defined below average as less than or equal to one standard deviation below the average teacher value-added results in the subject-subgroup combination. We assigned a value of average to all other teachers with eligible subgroup value-added results. We produced results for student subgroups based on limited English proficiency status, individualized education program status, and proficiency levels on the pre-test from the same content area as the post-test.

[13] Whereas we did not report full-sample value-added results for teachers with fewer than 10 students, we used a cutoff of seven students for subgroup value-added because it meant giving more teachers subgroup results. Although value-added results based on so few students will tend to be imprecise, this is less of a concern for the subgroup results than for the full-sample results, primarily because the subgroup results do not directly contribute to a teacher's TLE score.

REFERENCES

Arellano, Manuel. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics,* vol. 49, no. 4, November 1987, pp. 431–434.

Buonaccorsi, John P. *Measurement Error: Models, Methods, and Applications*. Boca Raton, FL: Chapman & Hall/CRC, 2010.

CTB/McGraw Hill. *Oklahoma School Testing Program, Oklahoma Core Curriculum Tests, Grades 3 to 8 Assessments, 2012–2013 Technical Report.* Monterey, CA: CTB/McGraw Hill, 2013a.

CTB/McGraw Hill. *Oklahoma School Testing Program, Oklahoma Core Curriculum Tests, End-of-Instruction Assessments, 2012–2013 Technical Report.* Monterey, CA: CTB/McGraw Hill, 2013b.

Herrmann, Mariesa, Elias Walsh, Eric Isenberg, and Alex Resch. "Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels." Washington, DC: Mathematica Policy Research, April 2013.

Hock, Heinrich, and Eric Isenberg. "Methods for Accounting for Co-Teaching in Value-Added Models." Washington, DC: Mathematica Policy Research, June 2012.

Liang, Kung-Yee, and Scott L. Zeger. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, vol. 73, no. 1, April 1986, pp. 13–22.

Morris, Carl N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association,* vol. 78, no. 381, 1983, pp. 47–55.

Walsh, Elias, Albert Y. Liu, and Dallas Dotter. "Measuring Teacher and School Value Added in Oklahoma, 2012–2013 School Year." Washington, DC: Mathematica Policy Research, May 2014.

This page has been left blank for double-sided copying.