Tennessee Comptroller of the Treasury

Offices of Research and Education Accountability

Justin P. Wilson, Comptroller

Legislative Brief

Use of Value-Added in Teacher Evaluations: Key Concepts and State Profiles

Linda Wesson
Associate Legislative Research Analyst

Kim Potts
Principal Legislative Research Analyst

Kelly Hill
Associate Legislative Research Analyst

March 2015

KEY POINTS

- Teacher evaluation systems in many states have undergone significant changes over the past four years in part as a response to federal initiatives, such as the Race to the Top and Teacher Incentive Fund grant competitions and No Child Left Behind waivers. Changes in evaluation systems are expected to continue, particularly among systems that include a component based on student test scores, as states adopt new standardized tests, and as the U.S. Department of Education revises its policies and timelines on the teacher evaluation component of NCLB waivers.

- The intent behind adding a student performance component to teacher evaluation systems was to broaden teacher accountability and focus on instructional practices that improve student outcomes. Measures of students' academic growth are often used in teacher evaluations because they indicate a student's academic progress over time rather than a student's proficiency level at a single point in time. Statistical growth models like TVAAS are designed to use student growth data to identify how well teachers help students achieve academic growth.

- Teacher evaluation systems vary considerably by state and can also vary by districts within a state. Among the most significant differences are:
    - statistical models selected for measuring student growth, and the grades and subjects for which the necessary standardized test data are available for such models,
    - methods used for measuring student growth for courses and grades without standardized tests,
    - determining how to combine multiple evaluation components into a single overall evaluation score or rating, including the weighting of components, the number of evaluation performance categories, and the criteria for inclusion in each category, and
    - levels of local control and flexibility allowed within the evaluation system.

- As states and districts have shifted to more frequent, structured, and rigorous teacher evaluation systems, policies that prescribe using evaluation rankings in teacher personnel decisions have also been changing significantly. States and districts are increasingly using teacher evaluation results for high-stakes decisions involving tenure, retention, compensation, and dismissal.

INTRODUCTION

This report was prepared by the Tennessee Comptroller's Offices of Research and Education Accountability (OREA) in response to a request by the General Assembly's House Education Committee. The report provides background information on seven questions related to incorporating statistical growth model (value-added) data into teacher evaluations (see pages 7-22):

1. Which teachers are required to have statistical growth model data as part of their evaluations?
2. What weight is given to statistical growth model data in teacher evaluations?
3. What student academic growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?
4. What components are part of teacher evaluations in addition to student growth data?
5. How are teacher evaluation performance ratings applied to personnel decisions?
6. What type of statistical growth models are used for teacher evaluation?
7. What issues and concerns are states and districts facing from their use of statistical growth model data as part of teacher evaluations?

These questions were used to build profiles of seven states and districts (including Tennessee) that are currently using or planning to use statistical growth models for teacher evaluation.

Exhibit 1: Summary of State and District Profiles

State and District Profiles	Statistical Growth Models Used	Required Weight in Teacher Evaluations (for teachers of tested classes)	Year of full Implementation	High Stakes Decisions that Factor in Teacher Evaluation Scores	Years of Teacher's Growth Estimates in Evaluation
Tennessee (pgs. 24-40)	TVAAS	35.0%	2011-12	tenure, dismissal, layoffs, (pay plans as a local option beginning 2014-15)	up to 3
Colorado (pgs. 41-44)	Growth Percentile Model	districts determine	2015-16	tenure	at least 2
District of Columbia (pgs. 45-48)	Mathematica model	35.0%	2009-10	pay plans, career path, dismissal	1
Florida (pgs. 49-54)	AIR model	50.0%	2015-16	dismissal, merit pay	up to 3
Los Angeles (pgs.55-59)	VARC model	to be determined (30.0% proposed)	to be determined	to be determined	up to 3
North Carolina (pgs. 60-65)	EVAAS	16.7% (OREA estimate)	2015-16	to be determined	3
Ohio (pgs. 66-71)	EVAAS	42.5%	2014-15	dismissal, promotion determined by districts	up to 3

State Trends

The use of statistical growth models (also known as value-added models) to measure teacher effectiveness has been increasing as states and school districts adopt new teacher accountability policies. Federal grants and waiver programs over the past five years have created incentives for states to evaluate teachers more systematically and to incorporate student growth measures into the evaluations.

While some states had been producing statistical growth model data before the federal incentives, in several cases their data was used only at the school and district levels, not at the individual teacher level.[a] In Tennessee, where individual teacher data had been calculated for years, its use was primarily diagnostic – to flag areas of concern or make better assignments of teachers to classrooms where they could be most effective – rather than evaluative.

As of 2013, 30 states required teacher evaluations to incorporate a student growth measure based on standardized state tests for tested grades and subjects.[1] States' teacher evaluation systems and their use of statistical growth models continue to change for a number of reasons:

- transitions to tests aligned with new state standards are expected to affect students' scores, and in some cases could affect the resulting growth model data and subsequent teacher evaluation ratings (see box "How is the transition to new standardized tests affecting teacher evaluations?"),
- longer than expected timelines for states and districts to develop, adopt, and implement new evaluation systems that incorporate statistical growth model data, and
- ongoing U.S. Department of Education policy decisions allowing states to delay implementation of No Child Left Behind waiver requirements for incorporating student performance data into teachers' evaluation scores and using evaluation scores in personnel decisions, as well as decisions on qualifying for waiver extensions and renewals.

Definitions

In this report, statistical growth models refer to various methods that:

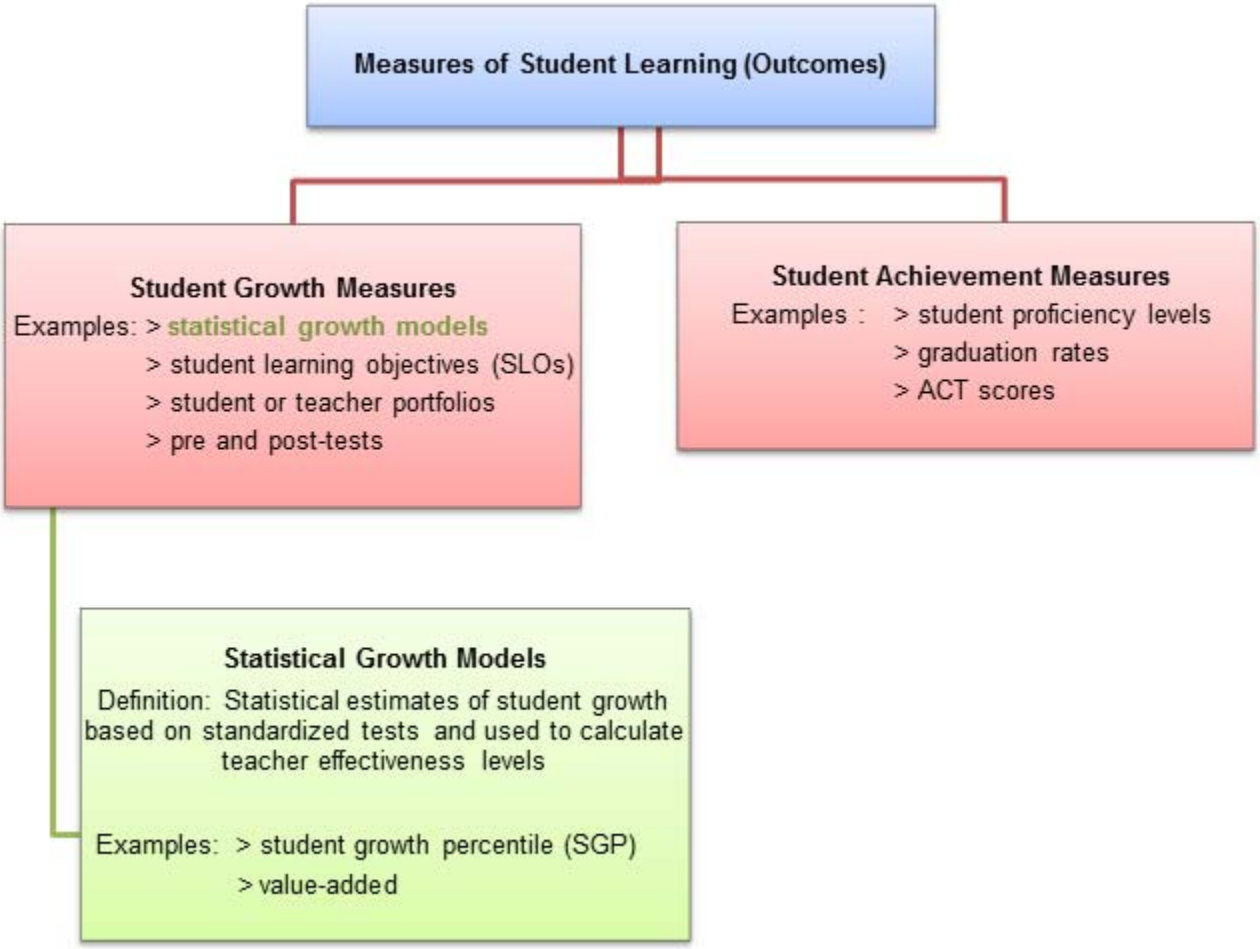
- apply statistical formulas to students' standardized state test scores to estimate students' academic growth during one school year in the context of their expected growth, and
- link students' academic growth to teacher performance by comparing students' test performance to performance estimates of comparable groups of students taught by other teachers.

This report uses the term "statistical growth model" to refer to the statistical models used by states and districts to calculate student growth estimates from standardized test data as a component of teacher evaluations. "Value-added" and "student growth percentile" are two types of statistical growth models, but **because "value-added" is more familiar and the common term in Tennessee, it is sometimes used interchangeably with the broader term "statistical growth model."** The report clearly distinguishes between value-added and student growth percentile models when necessary. (See more discussion at Question 6.)

[a] Examples of such states include Colorado, North Carolina, and Ohio.

Statistical growth model data is used for purposes other than teacher evaluation. States and districts may use such data for other education personnel evaluations, e.g., principals, and/or for professional development decisions. States, districts, and schools may use growth model data in accountability systems, as well as for school and district assessment and improvement efforts. In Tennessee, teachers are encouraged to use value-added data to identify effective teaching strategies for different student performance levels and to use individual student growth data to target instruction to meet students' needs.

Exhibit 2: Common Terms Used



Sources: Katherine E. Castellano and Andrew D. Ho, *A Practioner's Guide to Growth Models*, Feb. 2013, Council of Chief State School Officers, pp. 12-13, 21-22, http://scholar.harvard.edu/ (accessed Dec. 16, 2014); Stephen Raudenbush and Marshall Jean, *How Should Educators Interpret Value-added Scores*, Carnegie Knowledge Network, Oct. 15, 2012, pp. 2-3, http://www.carnegieknowledgenetwork.org/ (accessed Aug. 25, 2014); *American Statistical Association, ASA Statement on Using Value-added Models for Educational Assessment*, April 8, 2014, p.3, http://www.amstat.org/ (accessed April 9, 2014).

Background

Development of Growth Models: The No Child Left Behind Act (NCLB) passed in 2001 required states to identify failing schools based on their lack of ability to make "adequate yearly progress" (AYP).[2] The goal was to have all students meet an academic standard of "proficiency" as measured by newly-required annual tests on English and math in grades 3-8 and at least once in high school. Educators sought a way to recognize schools that were meeting learning goals (achieving growth) but were unable to meet AYP because their students started out further behind the "starting line" than other

students. In November 2005, the U.S. Department of Education (USDOE) announced a growth model pilot program under which selected states could use approved growth models in AYP determinations for their schools.[3] Tennessee was one of the first states selected to pilot growth models for AYP starting in the 2005-06 school year.[4] Some states, like Tennessee, had developed and had been using growth models earlier to identify districts, schools, and teachers that were more or less effective in improving students' standardized test scores.

In 2009, at the same time that growing numbers of schools were failing to meet AYP, a report by The New Teacher Project, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, gained wide attention for its finding that, across the school districts studied, 94 percent to 99 percent of teachers were evaluated as satisfactory or better.[5] The report found that differences in teacher effectiveness were rarely measured, recorded, or used to inform teacher employment decisions, like tenure, compensation, or layoffs.

Increased focus in education research on the links between student learning and teaching quality was reflected at the federal policy level in efforts "to strengthen accountability for teacher quality and . . . to define teacher effectiveness (at least in part) based on growth in student learning."[6] Other literature explored the need for evaluations that assess both the act of teaching (instructional practice) and the results of teaching (student learning).[7]

Development of New Teacher Evaluation Systems: In 2009, the USDOE announced the creation of its Race to the Top grant, which would award $4 billion to states through a competitive process.[8] In order to be eligible for the grant, states could not have legal, statutory, or regulatory barriers to linking data on student achievement or academic growth to teachers for the purpose of performance evaluations. One of the 19 application criteria required the states to demonstrate how they were "improving teacher and principal effectiveness based on performance."[9] This requirement was weighted more heavily than all but one other in the application scoring.[10] Specifically, state education agencies were to:

- establish clear approaches to measure student growth for each individual student;
- design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that differentiate effectiveness using multiple rating categories that take into account data on student growth;
- conduct annual evaluations of teachers and principals that include timely and constructive feedback on issues including student growth; and
- use these evaluations to inform decisions regarding professional development, compensation, promotion, retention, and tenure.[11]

Tennessee received one of the first Race to the Top grants in 2010 for $501 million.[12]

Another federal competitive grant program, the Teacher Incentive Fund (TIF), awarded grants to states, districts, and others for the development and implementation of teacher and principal performance-based pay programs, based primarily on increased student achievement to attract top talent to high needs schools. Applicants for the funds had to specify how teacher effectiveness would be quantified and how student data would be included. Created in 2006, TIF grew from appropriations of just under

$100 million in its first few years to $400 million in 2010.[13] The Tennessee Department of Education and selected districts received TIF grants in 2010 totaling $72 million over five years and in 2012 totaling $5.5 million for an initial two years.[14]

How is the transition to new standardized tests affecting teacher evaluations?

States that have recently adopted new standards and are now implementing new tests aligned to those standards face some expectations that student scores will drop, due to both the increased rigor of the tests and the drop that can occur when any new assessments are first administered. States like Kentucky and New York, which have already implemented tests aligned with new standards, have seen significant drops in student test scores.

Other states (e.g., Maryland) have also seen a drop in student test scores, attributed to the mismatch between their new standards, which are addressed in school curricula, and existing assessments still based on the state's previous standards. Some states (e.g., Illinois) have raised the difficulty level of their state assessments to prepare students and parents for the more rigorous tests yet to be implemented.

Where teacher evaluation ratings are partly dependent on students' scores, concerns have led some jurisdictions – like the District of Columbia and New York – to temporarily stop using student test scores in teacher evaluations or to suspend personnel consequences from low teacher evaluation ratings. Other states (e.g., Tennessee) have student growth measures that are dependent on test scores but can accommodate transitions to new tests without necessarily leading to lower growth scores.

In August 2014, the U.S. Department of Education announced that states with NCLB waivers can delay the requirement to use student test results in teacher evaluations until 2015-16.

Sources: Arne Duncan, "A Back-to-school conversation with teachers and school leaders," Smartblogs.com, Aug. 21, 2014, http://smartblogs.com/; Thomas Rains, "Standardized Testing in Alabama and Nationwide," A+ Education Partnership, Dec. 2013, http://www.aplusala.org/; Common Core Illinois, "Frequently Asked Questions," http://commoncoreil.org/; SAS Institute, Inc., *Misconceptions about Value-added Reporting in Tennessee*, 2012, http://www.tn.gov/education/; SAS Institute, Inc., *Ohio Department of Education: Technical Documentation of EVAAS Analyses*, 2014, http://education.ohio.gov/; various news articles from the *Baltimore Sun, Capital New York* (Albany), *Cleveland Plain Dealer, Education Week, New York Times, The State.com* (Columbia, S.C.), Aug. 2013 – Aug. 2014.

In 2011, USDOE established waivers providing state education agencies with flexibility on specific No Child Left Behind Act requirements in exchange for rigorous plans to improve educational outcomes. (For more on the NCLB waiver in Tennessee, see the Tennessee Comptroller's OREA report, *Tennessee's No Child Left Behind Waiver*.) To qualify for the waivers, states had to address four principles, including "supporting effective instruction and leadership."[15] Specifically, states had to develop and implement teacher evaluation systems that would:

- contribute to continual improvement of instruction;
- differentiate performance using at least three performance levels;
- include multiple valid **measures, including student growth**, in determining performance levels;
- evaluate teachers regularly;
- provide clear, timely, and useful feedback, including feedback that identifies needs and guides professional development; and
- inform personnel decisions.

As of November 2014, the USDOE had granted 33 states and the District of Columbia extensions of their approved NCLB waivers through the 2014-15 school year, and released procedures for states to renew their waivers.[16] Most states will be eligible to renew for up to three years, but some (including Tennessee) will be eligible to apply for an expedited renewal of up to four years.[17]

The USDOE has offered states the option to delay implementing certain provisions in their waivers. Recognizing implementation difficulties in some states, the department first offered states the option to delay using results from teacher evaluations in personnel decisions from the 2015-16 school year to 2016-17.[18] In August 2014, recognizing the convergence for several states of their switch in standardized testing programs at the same time as they were implementing new teacher evaluation systems dependent on the standardized tests, USDOE offered waiver states more flexibility, announcing that states could delay incorporating state standardized test data into teacher evaluations until 2015-16.[19] Tennessee is not planning to seek these waiver provision delays since the state has already established its teacher evaluation system to include student growth data; existing statutes require or allow evaluation results to be a factor in tenure and dismissal decisions; and Tennessee does not plan to change its standardized testing program until 2015-16.

Key issues related to understanding how states and districts use student growth models in teacher evaluations

This section explains the key issues used to build the profiles of the states and districts included in this report. OREA analysts reviewed only states and districts that are currently using or plan to use statistical models to estimate student growth for teacher evaluation. The following questions focus on key ways that states and districts may vary in the models chosen and their application.

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

Typically, less than 50 percent of all teachers in a state have growth model scores.[20] Because statistical growth models rely on student standardized test data as an essential input, only teachers in grades and subjects with standardized tests will have growth model data. Measuring growth requires a comparison of students at two points in time; thus states are often limited to growth measures for grades 4-8 in math and English language arts because those were the subjects and grades in which states have implemented annual standardized testing as required by NCLB.[21] (Third grade tests would serve as the base year against which to measure the first growth period in 4th grade.)

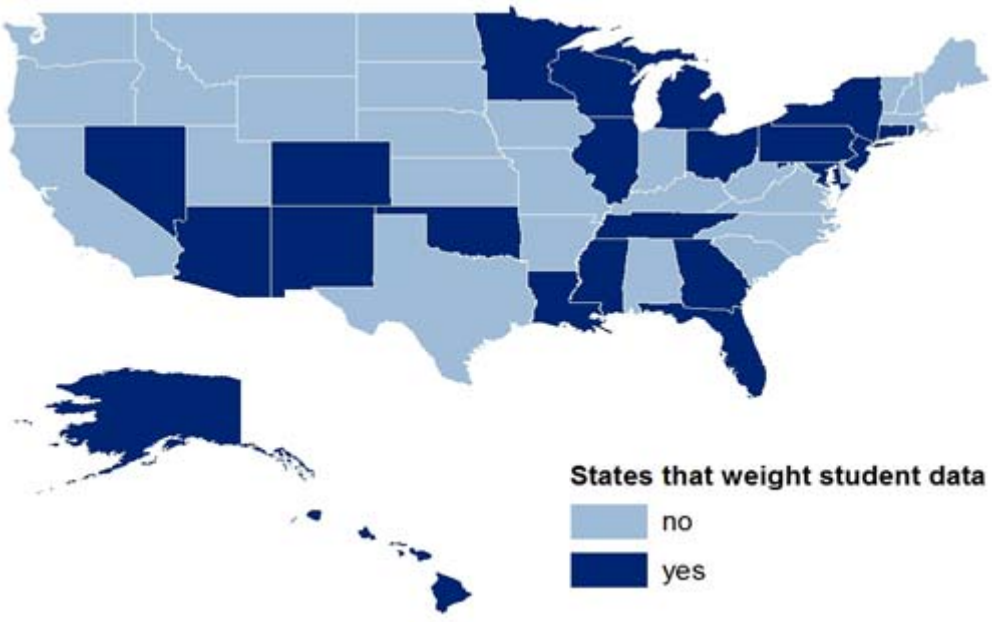
Although NCLB required at least one test during high school, students are assigned to high school courses in a way that requires different approaches for measuring growth. The subject matter at the secondary level does not necessarily follow a year-to-year content progression. Statistical models are available for calculating student growth based on standardized end-of-course tests administered in some high school courses, and states are adopting these tests in more subjects.

States have developed or are working to develop evaluation models for teachers who do not have growth model scores. (See more at Question 3.) Policymakers must consider how to combine evaluation components for teachers who teach some classes with and some without available growth model data, or who teach classes with students taking end of grade tests and other classes with students taking end-of-course tests.

2. What weight is given to statistical growth model data in teacher evaluations?

Twenty-three states assign explicit weighting to student achievement and growth in their teacher evaluation systems.[22] (See Exhibit 3.) Although there is variation in states' policies for weighting growth model data, a review of the literature did not find any state that required more than 50 percent of an evaluation to be statistical growth model data.[23] States with the lowest specifications were at the 15 percent to 20 percent level. Several states clustered at the 30 percent to 35 percent weighting. A three-year study by the Bill and Melinda Gates Foundation, the Measuring Effective Teaching Project, recommended a weighting of between 33 percent and 50 percent for state test score data to reduce year-to-year fluctuations and maintain the most accurate assessment of teachers' effects on student test scores.[24] Some states set a required level at which statistical growth model data can be weighted and some set a minimum, allowing districts to set their own specifications. Some states vary the percentage based on other factors. Some states may use a combination of individual teacher value-added scores and schoolwide value-added scores.

Exhibit 3: **States with Explicit Weighting of Student Growth and Achievement Components in their Teacher Evaluation Systems**



Source: Kathryn M Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 18, http://www.nctq.org/ (accessed April 15, 2014).

At least two jurisdictions reduced their weighting of growth model scores in 2014; the District of Columbia reduced the weighting of its growth model data in 2012-13.[b]

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

States have sometimes struggled to develop evaluation alternatives for teachers in grades and subjects that do not have standardized tests. In 2013, 18 states were identified as having specific policies on including student growth measures in evaluations of teachers in non-tested grades and subjects, compared to 29 states with policies on growth measures for teachers in tested grades and subjects.[25]

Approximately 60 percent to 70 percent of teachers nationwide teach in grades and subject areas for which there are no standardized tests.[26] Teachers most commonly in this group are those who teach grades K-2 and many high school classes (grades 9-12), as well as fine arts, physical education, world languages, career and technical, and special education. Some options states have used for teachers in non-tested grades and subjects include:

- substituting school or district level statistical growth model data when individual teacher data does not exist,
- weighting other evaluation components more heavily, and
- developing alternative measures of student growth not dependent on state standardized tests.

These alternate measures of student growth include:

- student learning objectives (see box "What are Student Learning Objectives?"),
- portfolios,
- subject-specific assessments, including commercially-developed standardized tests and locally-developed tests.

Tennessee's policy is that school level value-added data is to be the standard growth measure for teachers without individual data while other measures are in development.[27] The state has developed the alternative portfolio model for fine arts, physical education, and world languages and has a policy to continue developing alternative measures for more subject areas.

A teacher, or student, portfolio is a collection of evidence that demonstrates student growth over time. Portfolios generally include samples of teacher work (lesson plans, audio/visual recordings of classes, tests and grading rubrics, analysis of student learning data, and teacher reflections) and a representative sample of student work from throughout the year. The portfolios are evaluated by

[b] In 2014, Ohio passed legislation that reduces its weighting of value-added from 50 percent to 42.5 percent (a compromise between proposals of 50 percent and 35 percent). In June 2014, the Houston Public School Board of Education approved its superintendent's recommendation to reduce the weight of value-added scores from 50 percent to 30 percent for the 2014-15 school year based on research, discussions with the district's Advisory Committee, and the planned pilot of the new state evaluation system, which weights student growth at 20 percent. District of Columbia Public Schools revised its teacher evaluation system in 2012-13, reducing the weight of value-added from 50 percent to 35 percent. See more about Ohio and District of Columbia in this report's profiles.

experienced teachers in the same subject area through a blind peer review process. Student growth is calculated by examining evidence from two points in time during the school year.

States and districts are also developing or purchasing standardized tests from which statistical growth data can be generated (such as end-of-course tests), or substituting data from simple growth measures such as pre- and post-test results.

What are Student Learning Objectives?

Student Learning Objectives (SLOs) are academic goals that teachers and evaluators set for groups of students. Student learning objectives should be:

- specific and measurable,
- based on available prior student learning data,
- aligned to state standards, and
- based on growth and achievement.

SLOs may be developed to assess student progress within teachers' individual classes, within all classes across one grade or subject area, or for a subgroup of students (such as students below grade level) within multiple classrooms. Educators and their supervisors typically develop SLOs collaboratively at the beginning of the year, and design or select state or locally approved assessments to measure student growth targets.

Teachers' SLO scores are typically based on the proportion of students who have achieved the targeted goals by demonstrating growth or mastery on the selected assessment by the end of the year. Student learning objectives may also be referred to as student learning targets, student learning goals, and SMART (specific, measurable, achievable, results-oriented and relevant, and time-bound) goals.

In 2013, 14 states required student learning objectives to be part of teacher evaluations, and another six states permitted SLOs to be part of evaluations.

Sources: Natalie Lacireno-Paquet, Claire Morgan, Aniel Mello, *How states use student learning objectives in teacher evaluation systems: a review of state websites*, U.S. Department of Education, March 2014, http://ies.ed.gov/; Indiana Department of Education, *RISE Evaluation and Development System: Student Learning Objectives Handbook*, [no date], http://www.riseindiana.org/; Ohio Department of Education, *Ohio Teacher Evaluation System, 2014 Model*, Sept. 26, 2014, , http://education.ohio.gov/; Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 10, http://www.nctq.org/.

4. What components are part of teacher evaluations in addition to student growth data?

Types of Measures

Most education policymakers and researchers advocate for the use of multiple measures to evaluate teachers' effectiveness. In the past, teacher evaluations were often based solely on reviews of teachers' instructional practice, primarily through classroom observations. Components most likely to be part of teacher evaluations now are classroom observations of teacher practice, student growth and achievement (including statistical growth model data), student or parent surveys, and others, such as measures of professionalism or school support.

Classroom observations, the primary method for assessing instructional practice, are generally conducted by principals, assistant principals, or other school administrators, but can be done by fellow teachers, instructional specialists, or outside evaluators. Evaluators may also review lesson plans and other documentation related to instruction. Feedback for teachers on specific aspects of classroom

management and instructional techniques is one strength of the classroom observation component in improving teacher practice.

Classroom observations may be formal, with a structured scoring rubric, or may be conducted as informal walkthroughs, and may include review of lesson plans or other documentation. The duration of a classroom observation varies, ranging from a full day, a class period, or a few minutes and may be announced or unannounced. The number of classroom observations can vary by the teachers' years of service, type of license held, or previous evaluation ratings. Some teachers may be videotaped for evaluation purposes.

Student and parent surveys are a more recently developed tool for teacher evaluation. Student perception surveys ask students about specific teachers and classrooms and collect data on teachers' instructional practice, classroom management, and rapport with students. Studies find that student surveys are a valid and reliable measure in predicting student achievement gains, and that "students seem to know effective teaching when they experience it."[28] While student perception surveys are the most commonly used survey type, some states and districts have also begun to include parent surveys in teacher evaluations. As of 2013, 17 states required or allowed parent, student, or peer surveys to be included in teacher evaluations.[29]

How are student surveys used to evaluate teachers?

As of 2013, eight states required the use of student surveys in teacher evaluations; another four states allowed them. Two of the most commonly used student perception surveys include the Tripod Project survey and My Student Survey. Both surveys ask students to rate a number of statements on a five point scale of agreement, ranging from "totally true" to "totally untrue," or "never" to "every time."

The Tripod Survey consists of 80 questions in seven categories of teaching practice: caring, captivating, conferring, controlling, clarifying, challenging, and consolidating. The My Student Survey consists of 55 questions in six categories of teacher roles: presenter, manager, counselor, coach, motivator, and content expert.

Sample statements include:

- Students in this class treat the teacher with respect.
- My teacher explains difficult things clearly.
- My teacher notices when I am not participating in class.
- My teacher has us apply what we are learning to real-life situations.
- My teacher is able to answer students' questions about the subject.

Sources: Kathryn M Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 31, http://www.nctq.org/; Hanover Research, *Student Perception Surveys and Teacher Assessments*, Feb. 2013, http://dese.mo.gov/.

Classroom observations and student or parent survey results are sometimes referred to as the "qualitative" components of teacher evaluations to distinguish them from the "quantitative" components, which are based on student achievement and growth data, although often all components, both qualitative and quantitative, are converted to a numeric score for a teacher's overall evaluation score. (A detailed discussion of student growth data and statistical models can be found at Question 6.)

Student achievement measures, such as graduation rates, proficiency or passing rates on standardized tests (including state assessments, Advanced Placement, International Baccalaureate, and National Industry Certification assessments), and student scores on college aptitude tests, such as the ACT or SAT, may also be incorporated into teachers' evaluations.

Combining Multiple Measures

Once teacher evaluation components have been separately scored, states generally combine them to produce an overall evaluation rating. The method used to combine the scores, the number and alignment of categories, and the cut scores for each category impact teacher evaluation ratings and may introduce different types of classification or measurement errors. (See Question 6 for more about measurement error, a statistical factor different from "mistakes.") For example, some states, including Tennessee, convert all components to a number value on the same scale, such as 1-5, and weight them as prescribed in their evaluation systems. Some states weight different components on different scales (e.g., classroom observations are scored 1-5, but value-added is scored 1-3). Some states assign each component a performance level first (e.g., highly effective, effective), and then combine those in different ways.

In the past, state evaluation systems typically rated teachers as either satisfactory or unsatisfactory. As states redesign their evaluation systems, the number of states using more than two ratings has jumped from 17 in 2009 to 42 in 2011.[30] Today, most states use three to five rating categories for teacher performance; the most popular choice is four performance categories.[c] More categories help to differentiate among teachers' effectiveness but also increase the likelihood of measurement, estimation, and aggregation errors.

5. How are teacher evaluation performance ratings applied to personnel decisions?

The purposes for which teacher evaluations are used vary among states and districts (e.g., helping teachers improve performance, identifying struggling teachers for additional help or dismissal, and identifying top-performing teachers for retention through rewards or incentives).

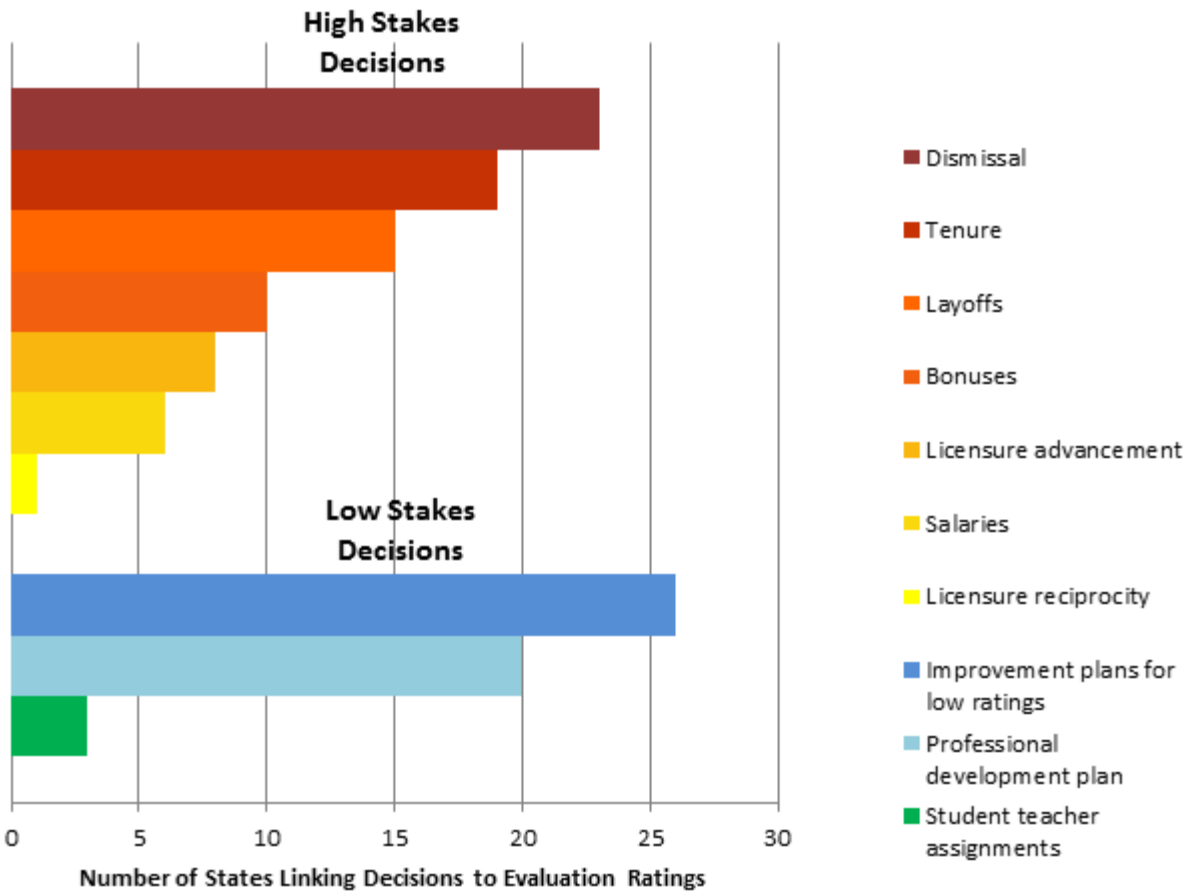
The extent to which teacher evaluation results are applied to personnel decisions can affect how teachers respond to their evaluation results, how teachers change their instructional practice, and, ultimately, how students learn. Personnel decisions can be classified as either low-stakes or high-stakes. Low-stakes decisions typically refer to those related to improving teacher practice, such as professional development, mentoring, future class assignments, and levels of required supervision or evaluation. For example, some states require more frequent classroom observations of teachers who had low teacher evaluation scores the previous year.

[c] Of the 42 states the National Council on Teacher Quality profiled as having policies on teacher evaluation performance categories, 31 have four categories (as does District of Columbia), eight states have three categories or require districts to have at least three, and three states, including Tennessee, require five categories. Kathryn M Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 34, http://www.nctq.org/ (accessed April 15, 2014).

High-stakes decisions typically have significant career impacts for teachers, including gaining or losing tenure, eligibility for pay increases or bonuses, career-path movement, layoff priority when positions are cut, and dismissal. The more evaluations are used in high-stakes decisions, the more attention teachers and their advocates usually pay to the details of the evaluation instruments and process.

Since the federal push to link teacher evaluation results with personnel decisions, states have increasingly used evaluations for high-stakes decisions. State policies have shifted significantly from past practices when tenure was usually granted to most teachers and pay schedules were based on years of employment and degrees earned. For example, the number of states where evaluations are tied to tenure grew from none in 2009 to 19 in 2013.[31]

Exhibit 4: Personnel Decisions Linked to Teacher Evaluation Performance Ratings



Source: Kathryn M Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. iv, http://www.nctq.org/ (accessed April 15, 2014).

6. What type of statistical growth models are used for teacher evaluation?

There are a variety of growth models available to states and districts, including:

- Student growth percentile models, also known as the Colorado Growth Model, by the National Center for the Improvement of Educational Assessment, originally developed by Dr. Damian Betebenner for Colorado,

- Education Value-Added Assessment System (EVAAS) and Tennessee Value-Added Assessment System (TVAAS) by SAS Institute, Inc., originally developed by Dr. William Sanders for Tennessee,[d]
- Value-Added Research Center (VARC) models,
- American Institutes for Research (AIR) models, and
- Mathematica Policy Research models.

Factors in states' and districts' choice of model may include the amount and type of data required, access to and reports of results, costs, trade-offs between clarity and accuracy, and intended use of the data. Common pricing inputs for value-added calculations, analysis, and data reporting include per student analyzed, per school in a district, and estimated provider staff time.

The purpose of using statistical growth model data in teacher evaluation is to help identify those teachers whose students gain more than a year's academic growth and those teachers whose students did not gain an average year of growth during the school year.

Differences among statistical growth models may stem from educational assumptions the developers make and the modeling strategies they choose. These decisions impact the types and amount of student data required, measurement error, and the reliability and validity of the results. The overview in this section highlights some of the key elements that models share as well as where they differ.

Statistical Models and Simple Growth Models

Simple growth models may track student performance from year to year (6th graders math scores in 2013 compared to 7th graders math scores in 2014) or from the beginning of a school year to the end (the same students' scores in the same year on a pre-test and post-test). These models are not designed to separate or isolate teachers' impact on student growth from non-teacher factors such as student motivation, family background (parents' hiring of tutors or providing enrichment learning experiences, for example), or school environment.

More complex growth measures like statistical growth (or value-added) models use statistical equations to control for outside factors in order to isolate the teacher's impact on students' growth and to estimate the amount of a teacher's impact on student growth compared to the average teacher's impact. Advanced statistical models generally include techniques for addressing missing data, measurement error within the standardized tests, changes in testing programs or test scoring scales and other factors (such as two teachers team-teaching in a classroom). These models report estimates within confidence intervals and with associated standard error.

Sources: Battelle for Kids, *Selecting Growth Measures: A Guide for Education Leaders*, 2011, http://www.edgrowthmeasures.org/; SAS Institute, Inc., *Misconceptions about Value-added Reporting in Tennessee*, 2012, http://www.tn.gov/education/; Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, et al., *Problems with the Use of Student Test Score to Evaluate Teachers*, Economic Policy Institute, 2010, http://www.epi.org/; Stephen Raudenbush and Marshall Jean, *How Should Educators Interpret Value-added Scores*, Carnegie Knowledge Network, Oct. 15, 2012, http://www.carnegieknowledgenetwork.org/.

[d] Dr. William Sanders of University of Tennessee originally developed TVAAS and Tennessee was the first state to collect value-added data. In 2000, Dr. William Sanders moved to SAS Institute, Inc., which bought the rights to TVAAS. The same basic model used in other states and districts is now called EVAAS.

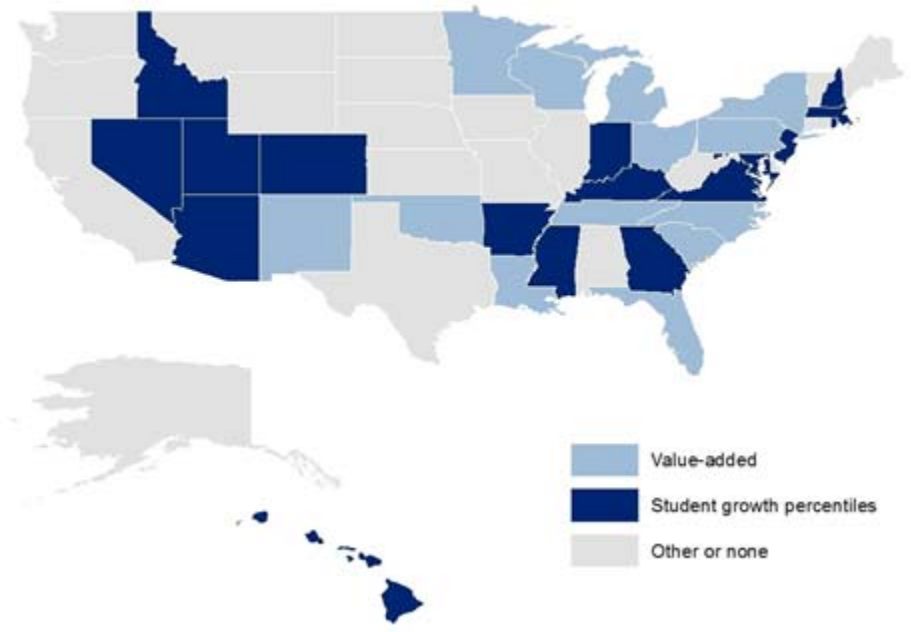
Differences Between Value-Added Models and Student Growth Percentile Models

Although the term *value-added* is often used to refer to all statistical growth models, education researchers have identified two general categories of statistical growth models: "student growth percentile" (SGP) models and "value-added" models.

- SGP models measure how much progress a student has made relative to other students based on their test scores.
- Value-added models attempt to isolate the impact a teacher has on students' test scores from other factors such as students' socioeconomic status.

State and federal trends to include growth measures in teacher evaluation scores have increased the use of SGP models, although such models were not originally designed to link students' academic growth with teachers' classroom performance as value-added models were. The data calculated and presented from SGP models is generally considered to be easier to understand than results from value-added models. One student growth percentile model is the Colorado Growth Model. See more about the Colorado Growth Model in the Colorado State Profile.

States Using Student Growth Percentile or Value-added Models for Teacher Evaluations



Source for map: Center for Great Teachers and Leaders, Database of State Teacher Evaluation Policies (Selecting Measures), as of Oct. 10, 2014, http://resource.tqsource.org/.

Sources: RAND Education, *Student Growth Percentiles 101, Using Relative Ranks in Student Test Scores to Help Measure Teaching Effectiveness*, not dated, http://www.rand.org/; Bruce D. Baker, Joseph Oluwole, and Preston C. Green, "The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era," *Education Policy Analysis,* Vol. 21, No. 5, Jan. 28, 2013, http://epaa.asu.edu/; Jim Hull, *Trends in Teacher Evaluation: How states are measuring teacher performance*, Center for Public Education, Oct. 2013, http://www.centerforpubliceducation.org/; Dan Goldhaber, Joe Walch, and Brian Gabele, *Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments*, Center for Education Data and Research, Aug. 10, 2012, http://www.cedr.us/; Elias Walsh and Eric Isenberg, *How Does a Value-added Model Compare to the Colorado Growth Model?*, Working Paper, Mathematica Policy Research, Oct. 2013, http://www.mathematica-mpr.com/.

Most statistical growth models share some common characteristics because they use equations based on data from students' standardized test results.

a. Scores are statistical estimates. The scores produced by the models are not measurements in the same way that a thermometer measures temperature, where the expectation is that all thermometers used at the same time in the same room should register the same temperature; statistical growth models instead produce best estimates within a range of possible scores. A teacher's growth model score is a statistical estimate based on the test scores of the students assigned to a teacher's classroom. The students assigned to a teacher's classroom constitute, in other words, a sample class of all potential classes that could be assigned to a teacher, and that classroom's test results are a sample of all possible classroom combinations of test results.

Research on student growth models finds that applying different growth models to the same student test results produces different estimates of teacher effectiveness.[32] Even when correlation among model results is considered high, different models can lead to teachers being ranked differently, depending on the number of performance categories a state has selected.[33]

How do value-added calculations for teachers differ from those for schools and districts in Tennessee?

In Tennessee, school and district level value-added calculations are based on all tested students, whereas individual teacher value-added includes only those students meeting attendance requirements.

Because of the smaller numbers of students in teachers' classrooms, the statistical formulas used to calculate value-added are somewhat different than those used for schools and districts. For instance, a statistical method known as "shrinkage" is applied in teachers' value-added calculations to help ensure conservative estimates, basically pulling their scores toward average until there is sufficient student data to indicate otherwise.

Other differences include providing for cases that teachers may be more effective in one subject than another or that teachers may share instructional responsibilities for a class.

Some differences between school and district value-added and teacher value-added apply only to grades 4-8, where students take end-of-year TCAP tests, because of the different model versions used for end-of-year versus end-of-course tests (see more at Tennessee Profile, Question 6). The formula for teacher level value-added provides for the accumulation of teacher effects over multiple years. Another difference is that the determination of school and district level value-added is based on calculating estimated average scores and subtracting the prior year from the current year (mean gains obtained from the differences in estimated means), whereas teacher value-added is based on calculating the differences in their effectiveness scores from the statewide average (deviations from the average gain).

Sources: S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, March 25, 2010, http://www.sas.com; Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014.

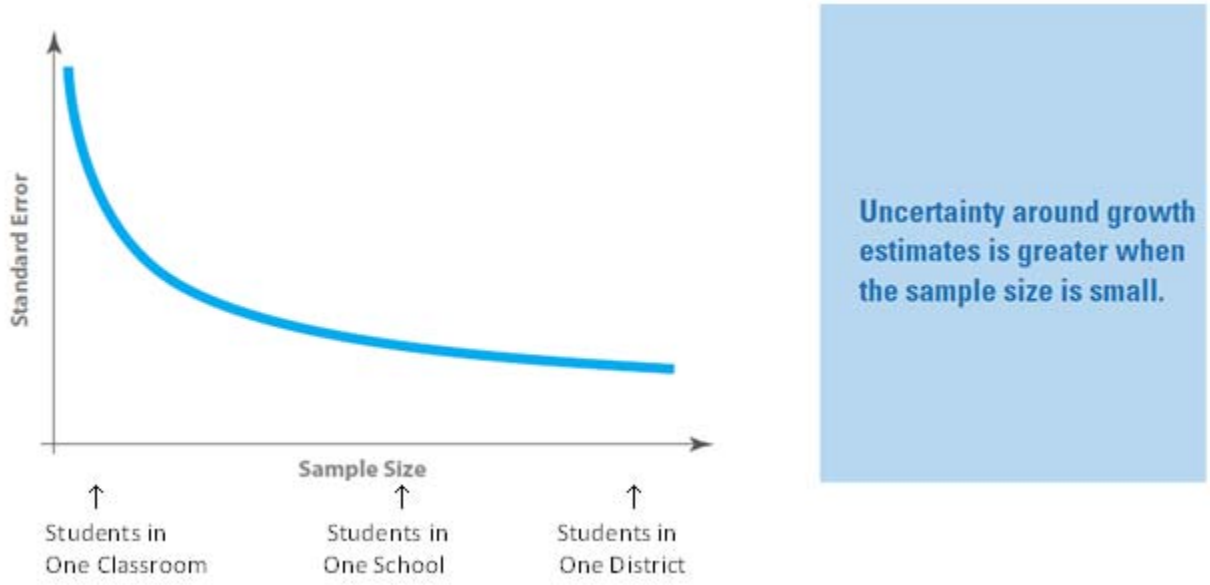
b. Statistical estimates include corresponding estimates of error. All statistical estimates have error rates; such errors are an expected element in statistics, not an error in the sense of a "mistake." This is analogous to opinion polls, which often have notations similar to "with an error rate of plus or minus 3 percent." This means the pollsters are reasonably certain that the scores they are reporting

from the sample of people surveyed are representative of the scores of the whole population, give or take 3 percentage points above or below the reported score.[e]

In the same way, estimates of teacher effectiveness levels might be expressed in terms of a reasonable certainty that a teacher's effectiveness is for example, a 1.5, with an error rate (also known as standard error) of plus or minus 0.7. This means that based on the sample of students assigned to that teacher's classroom that year, the best estimate of the teacher's "true" effectiveness is 1.5, but that if the teacher had had a different sample of students assigned, the best estimate could fall within the range of possible scores from .08 to 2.2 *(the 1.5 estimate minus 0.7 and 1.5 plus 0.7)*.

This range of possible scores is known as the confidence interval. States and districts use various strategies for considering confidence intervals when assigning teachers' performance level scores.[f] In general terms, the more points of data available to make an estimate, the lower the estimated error and the more precise the estimate. Thus, effectiveness estimates for teachers with particularly small class sizes are more prone to larger standard errors, as well as score fluctuations, from year to year.

Exhibit 5: Standard Error and Sample Size



Source: Adapted from Battelle for Kids, *Selecting Growth Measures: A Guide for Education Leaders*, 2011, p.12, http://www.edgrowthmeasures.org/ (accessed Dec 16, 2014).

Statistical growth models may use multiple approaches to adjust effectiveness estimates for teachers of small classes. One common method is to statistically "shrink" the growth estimates for teachers with few numbers of students back toward the average growth estimate for all teachers in a district or school. By this method, a teacher's growth estimate is not unduly influenced by a few outliers in student performance scores. Some states set a minimum number of students that must be linked to a

[e] The level of certainty is usually reported statistically – for example, that the pollsters are 90 percent or 95 percent confident that the reported scores are representative.

[f] For example, Tennessee divides value-added estimates by the standard error to produce a growth index, which is the data point used in teacher evaluations.

teacher before a value-added score is calculated.[g] Another method is to average several years of teachers' growth estimates. For example, Tennessee uses an average of up to three years of value-added data in teachers' evaluations, as the data becomes available.[34]

c. Growth model scores are based on standardized tests. Standardized tests are a commonly-used measure of student learning. Because standardized test results provide the raw data for statistical growth models, the quality of the standardized tests used and their measurement error have important implications for growth models.[35]

Tests are "estimates of student knowledge," based on a sample of possible questions.[36] How accurately a test measures students' learning depends on its *validity* – the extent to which the test measures what is being taught – and its *reliability* – how stable the test scores are across different versions of the same test (using different combinations of test questions) within a fixed time limit; hypothetically, the same student taking the tests repeatedly without additional learning experiences should earn similar scores. Measurement error is the difference between a student's actual knowledge level and the knowledge level measured by the test. Even well-designed tests have some level of measurement error (e.g., inexact alignment between test questions and academic standards, confusing wording, etc.). Measurement error can also result from students' mental and emotional state or conditions in the testing environment.

Of special importance when test results are used to measure student gains is whether the test is designed with sufficient *stretch* so that very low- and very high-performing students can be identified with questions that are easy or challenging enough to differentiate them from average students. Other test issues involve vertical scaling of scores (a methodology for taking results of tests from a series of grade levels and placing them on a common scale) and changes in test instruments that make student gains from one year to the next more difficult to calculate. Using different standardized tests as the input for one statistical growth model can produce different estimates of which teachers are most effective.[37] In general, higher quality standardized tests produce more valid and reliable indicators of student learning.

d. Growth model results exhibit year-to-year fluctuations. Although recent research suggests large swings in value-added scores are relatively rare, a number of studies have found that significant fluctuations in value-added scores can occur for a small percentage of teachers.[38] Using other performance measures in addition to value-added scores and using more than one year of teacher data are recommended as ways to mitigate these effects. Among the explanations for the year-to-year fluctuations are:

- Smaller sample sizes (number of students) in classrooms (compared to schools or districts) increase the standard error of the growth estimates.

[g] New York, for example, set the minimum number of students at 16 in order for a teacher to be provided a state-provided growth score. In Tennessee, to earn an individual growth score teachers must claim at least six full-time equivalent students who have at least three prior assessments (each subject test is counted separately).

- Teachers' performance may vary due to teacher-related factors (skills gained from training or experience, skills impaired from personal distractions) or due to outside factors, such as changes in school leadership or environment.
- Teachers may be more or less effective with a particular group of students, in a particular subject, or at a particular grade level. Principals' assignment of teachers to specific grades or courses and students to specific teachers and courses, as well as parental requests for specific teachers, may result in non-random assignment and may exacerbate this effect over time.
- Non-random assignment of students may also result in one year's classroom having more behaviorally disruptive students or a wider spread of student performance levels, making it harder for teachers to appropriately target instruction.
- Instruction that is not appropriately matched to each student's academic level within a classroom may contribute to greater fluctuations.

High-performing teachers tend to have more stable scores than low-performing teachers, and teachers' scores tend to stabilize after their first three to four years.[39] One analysis found that for teachers identified as ineffective after three years of teaching, "approximately half of them will continue to be identified as ineffective three years later."[40]

Some researchers have interpreted fluctuations in teacher scores as "low reliability of value-added measures."[41] SAS Institute, Inc., the EVAAS model provider, suggests that critics have confused "repeatability" of teacher value-added scores with "reliability," noting that "'perfect' repeatability is not the goal" and that some year-to-year variation is to be expected.[42]

In Tennessee, the Department of Education reports that the number of teachers whose value-added data fluctuated in one year from level 5 to level 1 was 0.4 percent and from level 1 to level 5 was 1.0 percent.[43] Tennessee teachers whose overall evaluation scores (value added + student achievement measure + classroom observation measure) fluctuated from the highest to the lowest rating totaled 0.1 percent, and from the lowest to highest rating was 3.8 percent, reflecting the same pattern of more rating shifts up than down.[44]

Despite their common characteristics, statistical growth models differ based on the underlying educational assumptions and the statistical and practical trade-offs made by their designers and users. The points of model differences discussed in this section generally have the potential to impact the uncertainty and standard error associated with any statistical estimate. (See parts (a) and (b) above.)

Number of prior years' test scores

Some models use results from only one or two prior years; others use up to five years of data to predict students' expected growth. Generally, more years of data improve the accuracy of growth estimates; however because models use prior years' test data differently, direct comparisons between models on this factor may not be reasonable. (For example, some models control directly for past achievement

using the test scores as a predictor variable. Others, such as TVAAS, use past achievement to refine estimates, using each accumulating year of test data to estimate students' gains more precisely.)

Missing data

Even if multiple years of student test data are used, some students will have incomplete or missing data due to mid-year transfers from another state or district, or absences on a testing day. Because there are fewer students in a single classroom than in the entire school, missing student data has a more significant effect on individual teacher growth model data than on school growth data. Some models exclude missing test records or students with missing data; others estimate missing data points for students; others, such as TVAAS, use the relationships between current and prior test scores for a group of students (e.g., a class) to estimate the group's average without estimating individual student data.

Controls for student demographics and other characteristics

"Controlling" for student characteristics refers to methods, which may be built into the process or applied during statistical analysis, to remove the effects of factors outside teachers' control in creating growth predictions. Some models include statistical controls for student demographic factors, such as gender, race, ethnicity, poverty, or for other student status characteristics, such as having disabilities, status as an English language learner, identification as gifted, attendance, mobility (changing schools or districts), or grade repetition.

Whether statistical growth models should control directly for student demographics and other characteristics is an issue much-discussed in the research literature.[45] The debate is over whether such factors are fully controlled for when using students' test score history (each student, in effect, acts as their own control, which is the approach taken by TVAAS) or whether explicit controls should be applied in addition to using past test scores to fully adjust for student differences (Florida's method, for example). Some models do not control for student factors in their calculations but may report the results for different student demographic groups separately. Tennessee teachers can review the academic growth for subgroups of their students when classified by demographic or other student characteristics.

Many statistical growth models control for student attendance. Tennessee, like some other states, prohibits the use of test score data in teacher value-added analysis if the tested student has not attended a minimum number of class days (150 per year).[46]

Other differences include whether models control for school effects (e.g., ratio of poor or special needs students) or classroom (peer) effects (e.g., class size, variance in student abilities). TVAAS does not include controls for school or classroom effects in its formula. Models may include procedures for attributing student gains to multiple teachers (e.g., when classrooms use a team-teaching model or when students transfer between classrooms or schools); the TVAAS model can calculate gains in shared teaching situations. Some models use formulas that allow each teacher's effects on student learning to accumulate over time, based on the premise that a student's current year progress depends

on learning acquired under previous teachers, in addition to the current teacher. These models can differ in how they calculate the persistence of past teachers' effects on student learning, either persisting without change into the future or diminishing over time (erosion). The TVAAS formula provides for past teachers' effects to accumulate over time without erosion.

7. What are the issues and concerns facing states and districts from their use of statistical growth model data as part of teacher evaluations?

Complexity and Communication

The technical nature and complexity of statistical growth models can make them difficult for non-statisticians to understand. Communicating information about complex growth model data to teachers and other stakeholders continues to be a challenge for most states and districts using these statistical models in teacher evaluations.

Alignment

As more states adopt evaluation policies that include multiple measures, and as the use of teacher evaluations in high-stakes personnel decisions has increased, questions have arisen among policymakers, educators, and researchers about which measures of teacher effectiveness are the best components for evaluation and to what extent such measures should align with each other. Research has found that the most commonly-used components do not necessarily show a strong relationship with each other. One study reports that relatively little is known about why components like student test scores and classroom observations by principals can be significantly different; the relationships between components may change as states continue to revise their evaluation systems.[47]

Research has found generally weak correlations between principal ratings of teachers and their value-added scores, but much stronger correlations when only high-performing teachers are considered.[48] Correlations are weak among factors that should be expected to correlate, such as value-added ratings and the amount and type of content covered in classrooms and the content in state assessments.[49] The Gates Foundation's Measures of Effective Teaching Project analyzed several classroom observation rubrics, value-added estimates, and student survey results, and found generally "low, positive correlations" among these measures.[50] The study also found that a combination of the three types of measures produced a more reliable measure that was better at predicting teachers' effectiveness.[51] (See more in the Tennessee Profile on state policies related to differences between classroom observation and value-added scores.)

To some, the mixed research results on correlation indicate a problem with one or more of the measures. The theory is that if measures are valid indicators of good teaching and a teacher is doing a good job, then different measures should all reflect good performance. To others, the mixed results confirm that teaching is a multi-faceted process and that multiple measures are needed to capture the full complement of teaching skills. Skills that principals and others may value in teachers include compassion, enthusiasm, fairness, the ability to work well with teams, and mentoring or other contributions to the school, which are not directly measured by students' performance on state

assessments. As one researcher explained,

> You can think of principal evaluations and value-added as measuring two different elements of 'quality instruction' in the same way that temperature and humidity are two key elements of 'quality weather.' Sometimes temperature is high and humidity is low, sometimes these are reversed, and sometimes they are similar. So, it's no surprise that principal evaluations differ from value-added [ratings].[52]

Lawsuits

As states and districts increase the use of teacher evaluation results for high-stakes personnel decisions, teachers are more likely to question the fairness, transparency, reliability, and weighting of evaluation components.

Teachers who believe they have not been fairly evaluated have filed lawsuits in some jurisdictions, including Tennessee, New Mexico, New York, and the Houston Independent School District.[53] A federal judge ruled in a Florida teacher lawsuit that the teacher evaluation system's use of value-added data, while unfair, was rational within the definitions of the law, and did not rise to the level of unconstitutionality. (See more in the Florida Profile.)

STATE AND DISTRICT PROFILES

The following states and districts were chosen to provide a diverse sample of jurisdictions that are using or planning to use statistical growth model data as part of their teacher evaluation systems. In addition to an in-depth profile of Tennessee's teacher evaluation system, summaries of selected jurisdictions are included for comparison purposes. Jurisdictions were selected based on several factors:

- Inclusion of at least one example from each major type of model or model provider, specifically the Colorado Growth Model (*Colorado*), SAS – EVAAS (*North Carolina, Ohio, Tennessee*), VARC (*Los Angeles*), Mathematica (*District of Columbia*), and AIR (*Florida*).
- Inclusion of some states considered by the U.S. Department of Education to be at the forefront of linking teacher evaluations to student growth (*District of Columbia, Florida*, *Tennessee*).[54]
- Inclusion of states that use a model similar to Tennessee's for comparisons in application (*North Carolina* and *Ohio* both use SAS-EVAAS).

States vary in the amount of flexibility they allow local districts in teacher evaluation design and implementation. These profiles describe only the states' recommended teacher evaluation system or the state framework that guides local district decisions.

TENNESSEE

Background

In January 2010, the Tennessee General Assembly passed the Tennessee First to the Top Act, which required:

- all teachers to undergo annual evaluations, with 50 percent of the evaluation based on student achievement measures, including at least 35 percent based on student growth data, and
- evaluations to be a factor in teacher personnel decisions, including promotion, retention, tenure, and compensation.[55]

Tennessee received a $501 million Race to the Top grant in July 2010.[56] In its application, the state reaffirmed its commitment to implementing the teacher evaluation reforms established by the First to the Top Act.

In June 2011, the State Board of Education adopted the Tennessee Educator Acceleration Model (TEAM) as the state model for the new teacher evaluation system. Districts across the state were required to adopt and implement the state model or adopt one of three approved alternate evaluation models by the 2011-12 school year.[h] In 2012, the Tennessee Department of Education (TDOE) permitted districts using the state evaluation model to make approved modifications to meet local needs through a process known as "TEAM Flexibility."[i]

Tennessee was granted a No Child Left Behind waiver in 2012. In its waiver proposal, the state again reaffirmed its commitment to its new teacher evaluation system. The U.S. Department of Education awarded Tennessee a one-year waiver extension in September 2014, and has notified the state of its eligibility for a longer waiver renewal.[57]

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

Tennessee's statistical growth model data is provided through the Tennessee Value Added Assessment System (TVAAS) for teachers in grades 4 through 8 in English language arts, math, science, and social studies. The data is compiled using student scores from the Tennessee Comprehensive Assessment Program (TCAP) *end-of-grade* tests. Value-added data is also calculated for high school teachers of English I, II, and III; Algebra I and II; Biology I, Chemistry I; and U.S. History, using student scores from TCAP *end-of-course* (EOC) assessments.[58] A statutory change in 2013 allows special education

[h] The three State Board-approved alternative models for teacher evaluation included (1) Project Coach, (2) TEM (Teacher Effectiveness Measure) and (3) TIGER (Teacher Instructional Growth for Effectiveness and Results).

[i] As of 2014, more than 70 districts have taken advantage of this increased flexibility. Flexible practices approved by TDOE include, but are not limited to, using multiple observers for observations, sequence of observations, introduction of additional unannounced visits, conducting video observations, and implementing evaluation-based mentoring programs and adopting student surveys as a component weighted at 5 percent. Tennessee Department of Education, *Teacher Evaluation in Tennessee: A Report on Year 2 Implementation*, 2013, pp. 16-17, http://www.tn.gov/education/ (accessed Sept. 11, 2014).

teachers to receive individual value-added data.[59] Districts have the option to conduct student testing in grades K-2, and value added-data is calculated for K-2 teachers in districts that choose to do so.

Teachers with individual TVAAS or alternative growth scores comprised approximately 47.4 percent of all classroom teachers in 2012-13 and 47.6 percent in 2013-14.[60] (See more about alternative growth scores at Question 3.)

2. What weight is given to statistical growth model data in teacher evaluations?

Student growth model data account for 35 percent of the overall evaluation score for teachers with individual TVAAS data or alternative growth model data.[61] (See more about alternative growth model data at Question 3.) Teachers who receive individual TVAAS scores of average or above (level 3, 4, or 5) may choose to have their TVAAS data weighted at 50 percent of their overall evaluation, and teachers who receive high individual TVAAS scores (level 4 or 5) may request that their TVAAS data count for 100 percent of their final evaluation score if allowed by their school district.[62] Teachers' value-added data from all the courses they teach with student growth results are combined and weighted proportionally to produce one composite value-added score for evaluation purposes.

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subject areas?

The student growth measure for teachers of most non-tested grades and subjects is the value-added data for the school or district in which they teach, although it is given a reduced weight of 25 percent in the overall evaluation score (instead of the 35 percent used for teachers with individual value-added data).[63] TDOE calculates four types of school-wide composite scores: (1) overall composite, (2) literacy composite, (3) math composite, and (4) math/literacy composite. Districts or schools select and assign the composite score most closely aligned to each teacher's responsibilities from an approved department list. (See Exhibit 1.) Teachers in alternative schools or itinerant teachers assigned to more than three schools use district-wide, rather than school-wide, composite scores.

Exhibit 1: Approved TVAAS Composites for Teachers in Non-tested Subjects

Subject	Recommended TVAAS Composite
Academic Intervention	Overall, Literacy, Math or Math/Literacy[(a)]
Computer Technology	Overall
Career and Technical Education	Overall, Literacy, Math or Math/Literacy[(b)]
English as a Second Language	Overall, Literacy
Fine Arts	Overall, Literacy
Health, Wellness, and Physical Education	Overall
High School Core Non-Tested	Overall, Literacy, Math or Math/Literacy
World Languages	Overall or Literacy
Early Grades	Overall or Math/Literacy

Notes: (a) The overall composite includes all tested areas, including science and social studies.
(b) Career and Technical Education (CTE) may use a narrowed version of a school-wide composite that includes only CTE students.
Source: Tennessee Department of Education.

TDOE has approved alternative models for measuring student growth in fine arts, world languages, and physical education. Districts may adopt these subject-specific models to provide teachers with individual student growth data. (See Exhibit 2.) The data from these alternatives is weighted at the same 35 percent as individual TVAAS data in teachers' overall evaluation scores.[64]

Exhibit 2: **Alternative Models to Measure Student Growth**

Subject	Number of Districts Using, 2013-14	Number of Districts Using, 2014-15 (expected)
Fine Arts	12	17-18
World Languages	1	1
Physical Education	(unavailable)	3

Source: Tennessee Department of Education

School-wide value-added data is expected to be used for a gradually declining number of teachers as the department continues to expand the number of alternatives for measuring student growth, according to State Board policy.[65]

4. What components are part of teacher evaluations in addition to student growth data?

All teachers' evaluations include a student achievement component weighted at 15 percent.[66] Student achievement measures are selected by the teacher and evaluator from a TDOE-approved list that includes state assessments (TCAP), school-wide TVAAS, graduation rates, and commonly used national assessments, such as ACT, SAT, Advanced Placement, or International Baccalaureate assessments.[67]

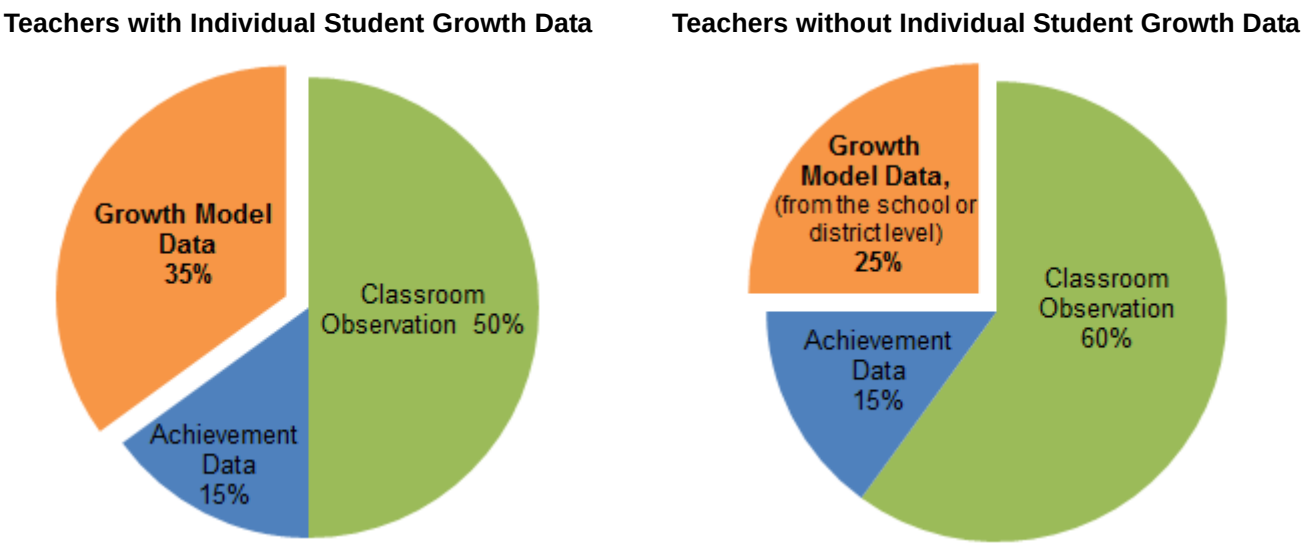
Under the Tennessee Educator Acceleration Model (TEAM) evaluation system, teachers are evaluated using a combination of student growth and achievement data weighted at 40 percent or 50 percent of the overall evaluation score, determined by whether they have individual student growth data or not.[68] Value-added data for teachers of tested subjects is weighted at 35 percent in their overall evaluations and for teachers of non-tested subjects at 25 percent.

The remaining portion of evaluations (50 percent or 60 percent) are required to include, but are not limited to:

- review of prior evaluations,
- personal conferences that include discussion of strengths, weaknesses, and remediation, and
- classroom observation(s) followed by a written assessment.[69]

This portion of the evaluation is focused on teachers' instructional practice, measured primarily through classroom observations using an approved appraisal form. Districts may opt to use student surveys as 5 percent of this component.

Exhibit 3: Tennessee's Components of Teacher Evaluation, 2013-14



Classroom Observations

To measure teachers' instructional practice during classroom observations, the State Board approved the observation instrument (or rubric) based on TAP, The System for Teacher and Student Advancement, an overall teacher quality and professional program provided through the National Institute for Excellence in Teaching.[70] The rubric scores teachers in four areas: designing and planning instruction, learning environment, instruction, and professionalism. Teachers' observation scores are based on a five-point scale. Districts may apply to use an observation instrument other than TAP, but any alternative must address the same four areas.

Classroom observations are conducted by principals, assistant principals, and other instructional leaders who have completed evaluation training and passed an evaluation certification test. In addition to classroom observation visits, evaluators are required to conduct a review of prior evaluations and hold personal conferences with each teacher to discuss strengths, weaknesses, and remediation. The frequency and length of classroom observations vary based on the teacher's level of experience and prior evaluation score. The number of annual observations ranges from three to six, a combination of formal and informal, announced and unannounced.[71]

5. How are teacher evaluation performance ratings applied to personnel decisions?

Tennessee differentiates teacher performance into five effectiveness levels:

Level 5 – significantly above expectations,
Level 4 – above expectations,
Level 3 – at expectations,
Level 2 – below expectations, and
Level 1 – significantly below expectations.[72]

State law requires evaluations to be a factor in teacher personnel decisions, including, but not limited to, promotion, tenure, dismissal, and compensation.[73] Tenure eligibility and, to a lesser degree, grounds for dismissal are personnel decisions that the state links specifically to teachers' evaluation scores. Other evaluation-based decisions are determined at the district level.

Licensure

In 2013, the State Board of Education adopted a new teacher licensure policy that, among other changes, set minimum overall evaluation scores and minimum value-added scores necessary for teachers to renew their professional licenses.[74] In 2014, the State Board of Education rescinded this portion of the new licensure policy, and the General Assembly passed legislation prohibiting the use of TVAAS data in teacher licensure decisions.[75]

Tenure

State law requires teachers to obtain an overall evaluation score of 4 or 5 during the last two years of a probationary period to be eligible for tenure.[76] Teachers who have received tenure remain under that status until they resign, retire, are dismissed, or are returned to probationary status.[77] Teachers who have received tenure status are returned to probationary status if, for two consecutive years of evaluations, they receive an evaluation score of 1 or 2, unless they received tenure prior to July 1, 2011, when the new tenure provisions became effective, and are grandfathered in.[78] After being placed on probationary status, teachers who have received two consecutive years of evaluation scores of 4 or 5 are again eligible for tenure, though the teacher may not continue working in the district if the director of schools does not recommend, or the local board of education does not grant, tenure.

Compensation

State law requires each school district to adopt a differentiated pay plan to aid in staffing hard-to-staff subject areas and schools, and in hiring and retaining highly qualified teachers.[79] Any differentiated pay plan based on performance must use teacher evaluation criteria approved by the State Board, such as classroom observations, value-added data, or other student achievement measures. The statutory requirement for districts to adopt a differentiated pay plan will be enforced starting in the 2014-15 school year.[80]

Dismissal or Suspension

According to state law, teachers may be dismissed or suspended for incompetence, inefficiency, neglect of duty, unprofessional conduct, or insubordination.[81] In 2011, the General Assembly expanded the definition of "inefficiency" to include low evaluation scores (level 1 or 2).[82] Consequently, teachers, including those who cannot lose tenure under grandfather provisions of the tenure law, may be dismissed or suspended for receiving low evaluation scores, although the law does not specify the number of low evaluation scores required to trigger eligibility for dismissal. Options for professional development or other supports for low-performing teachers are left to each district's discretion.

Transfers and Layoffs

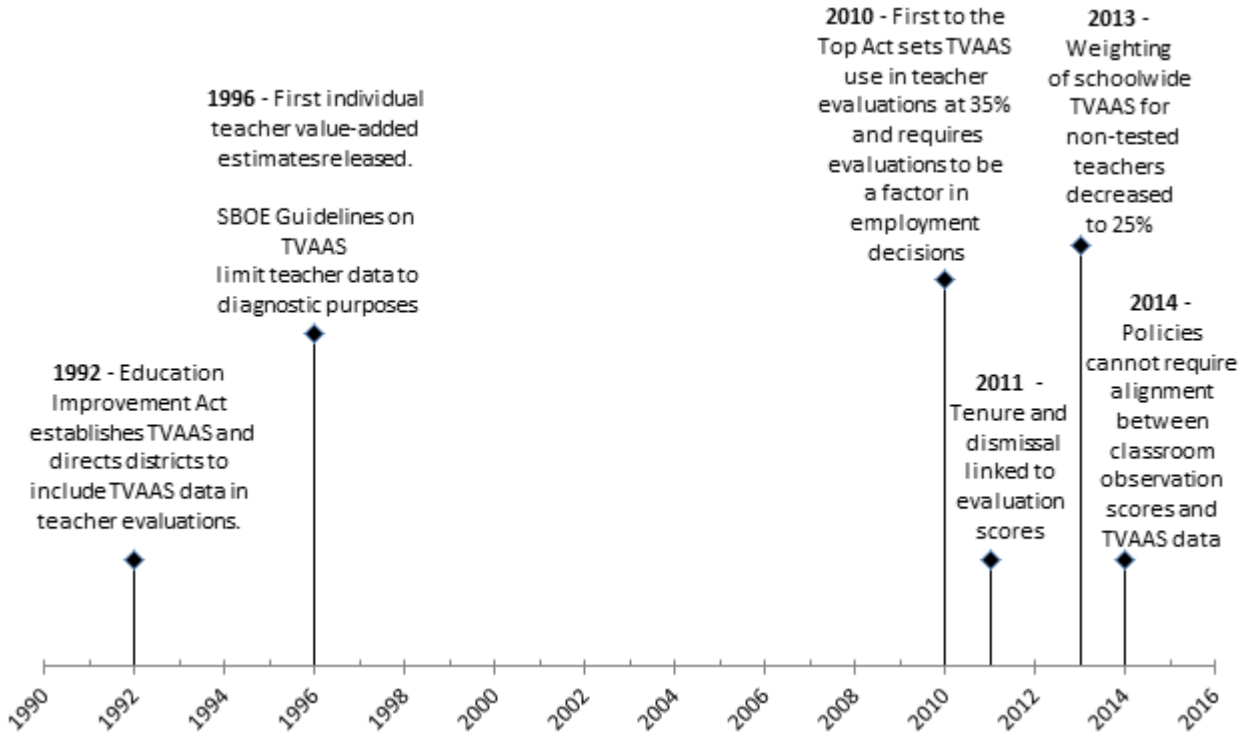
When a teacher transfers from one district to another, state law requires the teacher's former district to send the results from the teacher's past evaluations to the district into which the teacher is transferring. State law on layoffs (because of a reduction in positions, different from dismissal for cause) changed in 2014, allowing school boards to select teachers for layoffs on the basis of their overall evaluation ratings.[83] Districts are required to place on a list for reemployment any such laid off teachers who had earned an overall evaluation rating of level 3, 4, or 5, rather than those who had tenure status, as previously required.

Grievance Procedure

State law requires the State Board of Education to adopt grievance procedures by which evaluated teachers may challenge "only the accuracy of the data used in the evaluation and the adherence to the evaluation policies" at the local district level.[84] Grievances may be filed no later than 15 days after teachers receive the results for each component of the evaluation process, including the classroom observation, student growth measures (TVAAS or comparable data), and student achievement measures.[85]

The State Board's grievance procedure details a three-step process that includes the evaluator (step 1), the director of schools (superintendent) or designee (step 2), and the local board of education (step 3), which is the final step for grievances; local level grievance decisions may not be appealed to the state. The grievance procedure specifies that the final results of a teacher's evaluation may only be challenged if the teacher demonstrates by step 2 that procedural errors could "materially affect or compromise the integrity of the evaluation results."[86] Districts are responsible for ensuring that all teachers are aware of the grievance procedure.

Exhibit 4: How TVAAS Has Been Used in Teacher Evaluations



(See detailed descriptions in expanded timeline at Exhibit 8.)

6. What type of statistical growth model is used for teacher evaluation?

The Tennessee Value-Added Assessment System (TVAAS) was developed by Dr. William Sanders, who was originally with the University of Tennessee and later moved to SAS Institute, Inc. Tennessee first contracted for TVAAS in 1992, and continues to contract with SAS for value-added data at the district, school, and teacher level.[87] TVAAS data was used as part of the state's education accountability system *for districts and schools* until the federal No Child Left Behind accountability system replaced it. *Individual teacher value-added data*, initially used for identifying needed areas of improvement and creating teachers' professional development plans, became a state-required component of teacher evaluations in 2010.[88] (See Exhibit 4 for a brief timeline and Exhibit 8 for a detailed timeline on the use of value-added data in teacher evaluations.)

In general, TVAAS uses a longitudinally merged database of student test scores to estimate whether groups of students, on average, make more, less, or the same academic progress as students across the state.[89]

Computing Teachers' TVAAS Scores for Evaluations

Teachers' value-added scores represent the gains the students in their classes made relative to a state level benchmark. Individual students' scores may move up or down, but a teacher's growth score is based on whether the whole class, on average, makes more, less, or about the same progress as students across the state. Teachers' value-added data are incorporated into their evaluations as an average of up to three years of value-added scores (the most current one, two, or three years available for that teacher) for all eligible subject areas and grades.[90] A teacher's final value-added score is called the TVAAS evaluation composite.

Teachers' value-added data, including each single year score and multiple year averages, are reported with their standard errors, which vary in relation to the number of students in their classes, years of available data for the teacher, and the variance in students' scores. (See more about standard error in the Introduction, Question 6.) A value-added estimate (or effect score) divided by the standard error produces an *index score*, which is used to determine teachers' TVAAS levels on a five tier scale. (See Exhibit 5.) These five levels of effectiveness correspond to the five levels for teachers' overall evaluation rankings.

The state growth standard for index scores – the average expected academic growth for one year (also called the benchmark) – is always set at zero, although the actual state average for all teachers may rise above or fall below that standard. Using zero as the standard means that teachers whose students achieve more than one year's growth have positive value-added scores (above zero) and teachers whose students achieve less than one year's growth have negative scores (below zero). Teachers whose students achieve about one year's growth have scores clustered around zero (achieving the standard).

Individual teachers' value-added estimates cannot be compared because they may have different standard errors. Because the index scores include standard error, however, teachers' index scores can be compared from one teacher to another. Index scores are used for teacher evaluation purposes because they have factored in standard error.

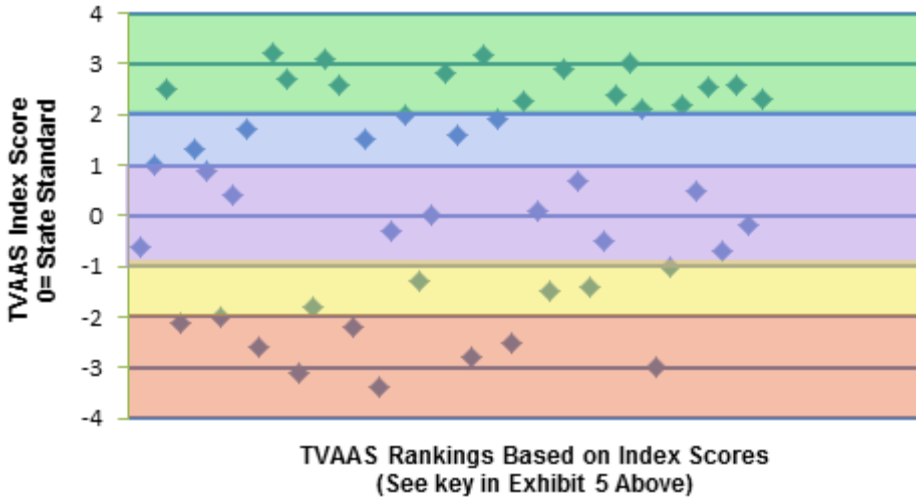
Exhibit 5: TVAAS Rankings by Index Scores

TVAAS Level		Comparison to State Growth Standard
5	**Most Effective**	***Index is 2 or more***: teachers whose students are making substantially more progress than the standard
4	**Above Average Effectiveness**	***Index is equal to or greater than 1 but less than 2***: teachers whose students are making more progress than the standard
3	**Average Effectiveness**	***Index is less than 1 but equal to or greater than -1***: teachers whose students are making the same progress as the standard
2	**Approaching Average Effectiveness**	***Index is less than -1 but equal to or greater than -2***: teachers whose students are making less progress than the standard
1	**Least Effective**	***Index is less than -2***: teachers whose students are making substantially less progress than the standard

Source: SAS Institute, Inc.

A sample of how Tennessee teachers with individual value-added scores might be distributed is at Exhibit 6. (For more in-depth data analysis of teachers' value-added and classroom observation scores, see the Tennessee Comptroller's Offices of Research and Education Accountability's report, *An Analysis of Tennessee Teacher Evaluation Data and Teacher Characteristics*.)

Exhibit 6: Hypothetical Example of Possible Teacher TVAAS Index Scores



Teachers who do not teach at least the minimum required numbers of students will not receive individual TVAAS scores.[j] The TVAAS model uses a *shrinkage* method to ensure conservative estimates of teachers' effects and reduce distortions of value-added measures for teachers with limited data due to small numbers of students. State law requires that the statistical model must be able to address situations in which two or more teachers share classroom duties.[91]

Students must be present for at least 150 days for year-long courses (about 83 percent of the standard 180 days) for their value-added data to be included in a teacher's score. Students must be present 75 days for courses taught on a block schedule (a semester-long course). An earlier provision in state law that disallowed any student eligible for special education services under federal law from being included in teachers' value-added scores was removed in 2013.[92] Tennessee law forbids teachers' TVAAS scores from becoming public record.[93]

Computing Student Growth used in Teachers' TVAAS Scores

Value-added was adopted into state law in 1992, and is described as follows:

> The statistical system shall have the capability of providing mixed model methodologies that provide for best linear unbiased prediction for the teacher, school, and school district effects on the educational progress of students.[94]

TVAAS is also described as a "multivariate longitudinal analysis" using "mixed model methodology."[95] This statistical model

- analyzes relationships between *multiple variables* at the same time (multivariate). Variables are any characteristics or quantities that can be measured or counted. Data points like student test scores, or a predicted amount of gain in test scores, are variables. Multivariate analysis is used when one or more variables, or outcomes, are influenced by one or more other variables, or predictors.

- analyzes the changes among a population (such as student test scores) over time (longitudinal).

- contains a *mix* of both fixed effects models and random effects models (mixed model). (Fixed effects and random effects models are types of statistical models. The choice of one over the other depends on characteristics of the variables used in the model and the inferences a researcher wants to make about those variables to a larger population. When mixed models are used to measure change within a group across time, they can adjust for missing student data, meaning students without complete data can still be included in the statistical analysis.)

[j] Tennessee teachers whose students take the end-of grade TCAP tests must have taught at least six full-time equivalent students before an individual growth score is calculated for them; teachers whose students take end-of-course tests must have taught at least 10 separate students, each with at least three previous assessments, who comprise six full-time equivalent students. (Full-time equivalency is based on the percentage of instructional time a teacher spends as the primary classroom instructor for the student.)

Originally established to estimate the impact of teachers, schools, and districts on student progress from grades 3 through 8 (with plans to develop high school subject value-added assessments), TVAAS now calculates value-added using two types of models:

- *multivariate response model* – used for the end-of-grade TCAP test results in grades 3-8, where the same testing program measures student achievement from year to year and test results are scored on a common scale, and

- *univariate response model* – used for end-of-course tests in high school and for assessments in grades K-2 for districts that use them.

The main differences between the two models are the *units of measurement* for estimating student growth, the *benchmarks* used as a state standard, and the *ability to re-estimate* each year's data.

Multivariate Model

The *multivariate model*, most similar to Dr. William Sanders' original TVAAS model, estimates a student's ranking among other students in *normal curve equivalent (NCE) units* which are similar to percentile rankings that are likely familiar to students and parents, except they are adjusted to be on an equal interval scale.[96] For example, an ACT composite score of 28 would put a student in the 90th national percentile, meaning that 90 percent of students who took the ACT scored 28 or below. Because of the typical bell curve shape of the distribution of student test scores, the differences between percentile ranks are not equal. A percentile rank of 30 is closer statistically to 40 on the bell curve than it is to 20. The use of NCE units addresses this issue by converting percentile rankings to a 0-100 scale with statistically equal intervals.[k]

Students' NCE rankings are based on their available prior year test results (up to five years).[97] Students' gains are the difference between their predicted NCE (percentile-like) ranking and their estimated NCE after they take their current grade test. The model uses all available data for students and includes records of students even if some of the data is missing, estimating means for data sets with missing data based on correlations of current and previous test scores.[98]

TVAAS Methodology Revisions Announced Jan. 27, 2015

The Department of Education announced that, beginning in the 2014-15 school year, it would make two revisions to the process for calculating value-added data for teachers in grades 4 through 8. Both changes affect the *multivariate* model. First, the *benchmark* or growth standard (the average expected academic growth for one year) will be based on students' *current year* performance (also known as *an intra-year benchmark*), rather than on a *base year.* Second, TDOE will no longer use the *re-estimation* process in calculating students' growth estimates.

Both of these changes align the methodology used for teachers in grades 4 through 8 more closely to that used for high school teachers (and K-2 teachers in districts that opt to participate) under the *univariate* model.

[k] A normal curve equivalent (NCE) was developed for the U.S. Department of Education by the RMC Research Corporation in 1976 to measure the effectiveness of the Title I Program of the Elementary and Secondary Education Act and is often used to measure gains over time.

Students' estimated gains are compared to a state benchmark to determine how their growth differed from an average year of growth for similar students. In the multivariate model a *base year benchmark* is used; currently the base year is 2008-09. Comparisons of student growth are made between the current year students and the 2008-09 students who started their school years with comparable prior achievement.[99] To calculate a teacher's classroom TVAAS score, the gains for each student in the class are averaged together.

Because the same test is given across grades 3-8, each year of additional test scores is accumulated to provide more precise estimates of each student's NCE (percentile-like) ranking. This feature of the TVAAS model means that students' NCE rankings for *previous years can be re-estimated using the current year's test score data*, resulting in re-estimation of teachers' value-added scores. The formula used to re-estimate all the past years' academic growth includes the accumulated effects of each student's teachers from past years. This is known as the *layering* of teachers' impacts on students' performance in each subsequent year. TVAAS assumes that teachers' impacts on students persist without diminishing or eroding over time.[100]

If a teacher's TVAAS score for a prior year changes in a subsequent year because of re-estimation, it does not affect the teacher's overall evaluation score already received for that prior year. The re-estimated score would be incorporated into the next value-added average (two or three years) and used as the student growth component in the next overall evaluation score.

Univariate Model

The *univariate model*, developed to measure student growth on end-of-course exams, compares students' past test results (a minimum of three test scores are required) to students with similar test histories to predict their expected scale scores.[101] *Scale scores* are the test results in the numeric format reported to students and parents. Students' academic growth is measured as the difference between their predicted and actual scale scores.

A student's estimated gains are compared to a benchmark to determine how the student's growth differed from the average growth of students similar to them. The univariate model uses an *intra-year benchmark*, meaning students are compared to all students statewide who took the same end-of-course exam in the current year. Students' gains are averaged for each class to produce the teacher's value-added score.

Because end-of-course tests are not given each year in the same subjects, the layering of teachers' impact on students and the re-estimation of past years' test scores are not part of the univariate model formula.[102]

Other Features of TVAAS

The TVAAS model does not directly control for student demographics or other student characteristics. Value-added data is calculated on the basis that student characteristics have been captured within all the available years of past test data; each student is, in effect, his or her own control. TVAAS does not

control for classroom (peer) effects or school effects. Tennessee's model does account for the possibility that a teacher may have varying levels of effectiveness in different subject areas and grade levels.[103]

Costs

Tennessee has contracted for TVAAS through a sole source contract since 1992, first with the University of Tennessee, and then, beginning in 2000, with SAS. At the time TVAAS was implemented, value-added was a new approach that had not been used before. The law directing the use of value-added was written to require the methodologies provided by Dr. William Sanders' model.[104]

Exhibit 7 shows the cost of Tennessee's TVAAS contracts since 1992. Services provided under the contract have expanded over the years to include enhanced reporting, website development, and additional data compilation, data analysis, and calculations. The cost increase since 2011 reflects additional duties to meet Race to the Top requirements, including the development of training for teachers, principals, and district administrators on how to use value-added information to improve student achievement.

Exhibit 7: Tennessee Contracts for TVAAS

Contract Period	Contract Party	Estimated Annual Average	Total Contract For Entire Period
July 1992 – Dec. 1999	University of TN	$306,731	$2,300,484
Jan. 2000 – Dec. 2003	SAS	$691,000	$2,764,600
Jan. 2004 – Dec. 2004	SAS	$1,130,000	$1,130,000
Jan. 2005 – Dec. 2010[(a)]	SAS	$1,500,000	$9,019,249
Jan. 2011 – Dec. 2015[(b)]	SAS	$2,800,000	$13,879,000

Notes: (a) Slightly more than $2 million of the total contract in this period was federally funded.
(b) $5.7 million of the total contract in this period was federally funded.

Sources: Office of Education Accountability, *The Education Improvement Act: A Progress Report*, Tennessee Comptroller of the Treasury, April 2004, pp. C-63 to C-64, http://www.comptroller.tn.gov/; Offices of Research and Education Accountability, *Tennessee Value-Added Assessment System: A Brief Overview*, Tennessee Comptroller of the Treasury, Aug. 2013, p. 2, http://www.comptroller.tn.gov/; Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014.

7. What issues and concerns does Tennessee face from its use of statistical growth model data as part of teacher evaluations?

Tennessee has faced issues of communicating its complex value-added model with teachers, as described in the introduction of this report. The state has also dealt with the specific issues of alignment and lawsuits.

Alignment

In 2014, the General Assembly passed Public Chapter 796, which amended the law at the request of the Tennessee Department of Education to ensure that there was no forced correlation of observation scores to match TVAAS scores.[105] TDOE officials were not aware that any such correlation was occurring, but understood that teachers had concerns that it could occur. The law states:

> No rules, policies, or guidelines shall be established that require the classroom . . . observation results pursuant to subdivision (d)(2)(C) to be aligned with TVAAS data.[106]

At the time, the department noted that the State Board had a policy about alignment, but it was to identify districts with large discrepancies and provide them with additional training.[107]

State Board policy states:

> . . . performance level discrepancies between individual student achievement growth scores and observation scores of three or more will be considered outside the acceptable range of results.
>
> The 10 percent of schools with the highest percentage of teachers falling outside the acceptable range of results will be required to participate in additional training and support as determined by the department. Districts that have 20 percent or more of their teachers fall outside the acceptable range of results will, as determined by the commissioner, lose their ability to apply for or implement alternate evaluation models or TEAM Flexibility the following school year.[108]

The Tennessee Department of Education states that there is not a conflict between the new law and the State Board policy because the policy does not "require alignment," but rather is used as an indicator to "assess whether additional support is needed."[109] Approximately 12 percent of teachers had gaps of three or more points between their TVAAS scores and their classroom observation scores in 2011-12, with slightly fewer (11.5 percent) having such gaps in 2012-13.[110] In 98 percent of these cases, the TVAAS scores were lower than the observation scores.

According to TDOE, 104 schools (5.9 percent) in 2012-13 were identified as having significant percentages of teachers with gaps of three or more between their TVAAS and observation scores.[111] Of those, 70 received some type of intensive support, which might include training for all classroom observers, department staff co-observing with school observers, or other support. Thirteen districts have been flagged based on 2012-13 data as having more than 20 percent of teachers with a three-point or more discrepancy between their TVAAS and classroom observation components.[112] The department had not imposed reductions of flexibility on those districts as of October 2014.

Lawsuits

In 2014, the Tennessee Education Association filed two lawsuits in federal district court on behalf of two Knox County teachers.[113] The suits were related to the way the teachers' TVAAS evaluation scores were used in the district's performance pay eligibility calculations; Knox County has tied compensation to teacher evaluations. As of November 2014, hearings in the cases had not been scheduled and the court had not made a final ruling.

Exhibit 8: Detailed Timeline —Teacher Evaluation and Value-added (TVAAS)

1982 Tennessee Department of Education (TDOE) commissions University of Tennessee professor Dr. William Sanders to develop the Tennessee Value Added Assessment System (TVAAS).

1991 TDOE begins contracting with the University of Tennessee for the production of TVAAS data.

1992 Public Acts, 1992, Chapter 535, the Education Improvement Act (EIA):

- establishes a new funding formula for public schools (Basic Education Program or BEP),
- creates a new local governance structure (school board appointment of superintendents, who are vested with all district personnel matters), and
- enacts an accountability system based on
  - expanded annual state assessments for students,
  - the TVAAS method for reviewing teacher, school, and district impacts on student learning,
  - performance goals for schools and districts, including sanctions, such as school or district probation and removal of local school officials, for failure to meet such goals, and
  - enhanced public reporting of school performance.

The EIA requires TVAAS data for teachers to be produced by July 1, 1995, for teachers with TCAP tests in grades 3-8. Students must have attended at least 150 days of the 180 day school year to be included in teachers' TVAAS calculations. Students eligible for special education services under the federal definition are excluded from teachers' TVAAS calculations. TVAAS data were not to be used "as a part of formal personnel evaluation" until three years of data were available and were not to be a public record.

The EIA also requires the State Board of Education to develop guidelines for districts' teacher evaluations that include classroom observation with written assessment, review of prior evaluations, personal conferences, and "other appropriate criteria including the Sanders model, related to the responsibilities of the employee."

1994 TDOE releases the first school level value-added data.

1996 TDOE releases the first teacher value-added data to local school districts.

The State Board of Education adopts guidelines for use of TVAAS data stating that teacher value-added data should be "limited to an advisory role until studies regarding its most appropriate use in teacher evaluation are completed." [(a)]

2000 TDOE implements a new teacher evaluation system statewide after three years of pilot testing. The Framework for Evaluation and Professional Growth (FEPG) rubric includes six domains that are evaluated through various components including classroom observations, teachers' self-assessments, reflection records, educator information records, and professional development/growth plans. Existing law required teachers with a professional license to be evaluated twice within a ten year period.

Individual teacher TVAAS data, if available, is to be part of the FEPG educator information record, and may be used to develop a teacher's required professional development plan.

TDOE begins contracting with SAS Institute for TVAAS, following Dr. William Sanders' employment with the SAS Institute.

2001 Congress passes the No Child Left Behind Act (NCLB), which requires the use of criterion-referenced tests, instead of (or in addition to) states' norm-referenced tests. The TVAAS model at the time was based on norm-referenced items from TCAP and used national norm gains as performance benchmarks. (In 2004, TDOE and Dr. William Sanders would enter discussions on converting TVAAS to use criterion-referenced test data instead of norm-referenced data.)

2004 The Tennessee House Education Committee holds a hearing on the use of TVAAS in response to a bill filed that would delete it from state law (HB 2270, 2004).

Public Acts, 2004, Chapter 928 revises school and district accountability criteria in order to align Tennessee's accountability system with federal No Child Left Behind standards. The revised criteria requires schools and districts to meet attendance and graduation rate targets, as well as both the

- federal criterion of adequate yearly progress in the percentage of students achieving proficiency in core subjects as determined by the State Board, and the
- state criterion of academic growth that is greater than or equal to the state growth standard for each subject and grade, as set by the commissioner with approval of the State Board, and that is measured by TVAAS.

2007 Public Acts, 2007, Chapter 376 increases evaluation requirements for professionally licensed teachers to one formal evaluation and two informal evaluations (assessments) every five years.

2009 The U.S. Department of Education (USDOE) announces competitive Race to the Top grants, which require the development and implementation of new teacher and principal evaluation systems that include student achievement as a component.

2010 Public Acts, 2010, First Extraordinary Session, Chapter 2, the First to the Top Act, establishes parameters of a new teacher evaluation system and creates the Teacher Evaluation Advisory Committee (TEAC). The law requires:

- 35 percent of teacher evaluation scores to be based on student growth data, either TVAAS or a comparable measure if the teacher does not have individual TVAAS data, and 15 percent to be based on other student achievement measures,
- annual evaluations of all teachers, which are to be a factor in employment decisions, including promotion, retention, tenure and compensation, and
- teacher value-added data to be available for sharing with approved teacher preparation programs if the data does not personally identify particular teachers.

Additionally, the act removes the requirement that "a teacher's effect on the educational progress of students may not be used as a part of formal personnel evaluation until data from three complete academic years are obtained."

The USDOE awards Tennessee a Race to the Top grant of $501 million.

TEAC begins developing teacher evaluation model guidelines and criteria, including the number and structure of classroom observations, the observation rubric to be used by evaluators, and the student achievement measures to be used for the 15 percent student achievement component.

2011 Public Acts, 2011, Chapter 70:

- revises tenure provisions, adding a requirement for two consecutive years of high overall evaluation scores (above or significantly above expectations, i.e., level 4 or 5) for eligibility, and providing for the loss of tenure after two consecutive years of low evaluation scores (below or significantly below expectations, i.e., level 1 or 2), and
- revises the definition of inefficiency – one of the statutory reasons for which a teacher may be dismissed – to include overall evaluation rankings of below or significantly below expectations.

The State Board adopts a new state model for teacher evaluation, the Teacher Educator Acceleration Model (TEAM), plus three alternative models proposed by some districts. The TEAM model includes:

- annual evaluations that differentiate teacher performance into five categories,
- a 35 percent student growth component, which is either the teacher's individual TVAAS score or, for teachers without individual growth data, the school-wide TVAAS score in lieu of comparable growth measures, to be developed and approved by TDOE,
- a provision allowing teachers with individual TVAAS scores in the top three levels to use their TVAAS score for 50 percent of their overall evaluation, replacing the 15 percent student achievement component, and
- a requirement of at least four observations per year, each with written feedback and a conference, to evaluate teacher instructional practice.[(b)] The TAP rubric is adopted as the statewide classroom observation rubric for teacher evaluations.

2012 The USDOE approves Tennessee's request for a waiver from specific NCLB requirements (also known as ESEA Flexibility). Waiver approvals hinge on states' commitments to four key principles, including implementation of teacher evaluation systems that meet federal guidelines.

All teachers receive evaluation scores under the state's new system.

TDOE changes the basis of school-wide TVAAS scores for teachers without individual data from a three-year average score to a one-year score and revises classroom observation requirements based on teachers' value-added and overall evaluation scores.

Public Acts, 2012, Chapter 811 makes the results of individual teacher evaluations exempt from public inspection.

2013 Public Acts, 2013, Chapter 105 makes several changes to how the teacher evaluation system incorporates TVAAS data, including:

- changing the weighting of the student growth data in overall evaluations from 35 percent to 25 percent for teachers that lack individual TVAAS data and use school-wide TVAAS data instead,
- allowing teachers with high individual TVAAS scores (level 4 or 5) to use this score for 100 percent of their overall evaluation at the discretion of the district, and
- deleting the provision that prevented inclusion of students who receive special education services in calculations of teachers' individual TVAAS data.

Public Acts, 2013, Chapter 369 allows local school boards to lay off teachers, when necessary due to a reduction in teaching positions or other reasons, based on their overall evaluation rankings and requiring that teachers affected by such layoffs who had received overall evaluation rankings in the top three levels be placed on a reemployment list. (The law became effective July 1, 2014.)

TDOE approves new measures of student growth for teachers of World Languages and Fine Arts to replace the use of school-wide TVAAS scores.

2014 Public Acts, 2014, Chapter 740 allows teachers scoring "significantly above expectations" (level 5) on each of their last three evaluations to request a waiver of any license renewal requirement.

Public Acts, 2014, Chapter 746 prohibits TDOE from revoking or non-renewing a teacher's license based on student growth data as represented by TVAAS, or some other comparable measure of student growth if no such TVAAS data is available. The legislation also removes the State Board's authority to set policies for the *revocation* of licenses and certificates after July 1, 2015, although the board retains its authority to set policies for *awarding* licenses and certificates.

Public Acts 2014, Chapter 796 prohibits the establishment of rules, policies, or guidelines that require teachers' classroom observation results to be aligned with TVAAS data.

Public Acts 2014, Chapter 885 allows teachers, rather than evaluators, to select the student achievement measures that represent 15 percent of their evaluations in cases where the teacher and evaluator cannot agree on the measure to be used, with verification by TDOE that the selected measure is appropriate to a teacher's assignments.

Notes: (a) Although this SBOE guideline is still listed as in force, TDOE staff indicate it was updated by State Board Policy 5.201.
(b) Apprentice teachers must have six evaluations.

Sources: **1982** - Ashley Ball, "Classroom Chronicles: The Man Behind the Numbers," Tennessee Department of Education, Oct. 7, 2013, http://tnclassroomchronicles.org/;
1991 - Office of Education Accountability, *The Education Improvement Act: A Progress Report*, Tennessee Comptroller of the Treasury, April 2004, p. C-63, http://www.comptroller.tn.gov/;
1992 - *Tennessee Public Acts*, 97th General Assembly, 1992, Chapter 535, An Act to enact the Education Improvement Act of 1991, and to amend and repeal provisions of Tennessee code Annotated, Titles 8 and 49, accordingly, pp. 39-40;
1996 - Guidelines of State Board of Education, 3.402, Guidelines for Use of TVAAS Data, adopted July 12, 1996, http://www.tn.gov/sbe;
2001 - Office of Education Accountability, *The Education Improvement Act: A Progress Report*, Tennessee Comptroller of the Treasury, April 2004, p. D-6, http://www.comptroller.tn.gov/ and Consortium for Policy Research in Education, *Assessment and Accountability in the Fifty States: 1999-2000*, Tennessee, July 2000, pp. 2, 13, 19, 20, http://www.cpre.org/;
2004 - Research Division of the Tennessee House of Representatives, *The Research Review*, Session Edition VII, March 4, 2004, pp.6-7, http://www.capitol.tn.gov/ and 103rd Tennessee General Assembly, Public Acts 2004, Chapter 928, AN ACT to amend Tennessee Code Annotated, Title 49, Chapter 1, Part 6, relative to school systems on probation, http://www.tn.gov/;
2010 - Deleted language in quotes was from the Education Improvement Act, Public Acts, 1992, Chapter 535, Section 4 (g)(4)(C); Joseph C. Conaty, Interim Director, Race to the Top, U.S. Department of Education, letter, addressed to Gov. Phil Bredesen, July 26, 2010, http://www2.ed.gov/;
2011 - Tennessee State Board of Education, Agenda Item III A, Educator Evaluation Models, June 14, 2011, http://www.state.tn.us/sbe/ and Policies of the State Board of Education, 5.201, Teacher and Principal Evaluation Policy, revised Nov. 4, 2011, http://www.tn.gov/sbe;
2012 - Tennessee Department of Education, Teacher Evaluation in Tennessee: A Report on Year 2 Implementation, 2013, pp. 9-10, http://www.tn.gov/education/.

COLORADO

Background

In 2010, the Colorado General Assembly passed legislation making substantial changes to the state's teacher evaluation system, including a requirement that local school boards, by 2013-14, base at least 50 percent of an educator's annual evaluation on student academic growth.[114] According to Colorado Board of Education rules adopted following the 2010 law, growth measures used in teacher evaluations must include results, if available, from the Colorado Growth Model (CGM), a statistical tool that measures student achievement and growth on statewide assessments each year and student academic growth over time. Districts may choose to include results from other state assessments, district assessments, and teacher-developed assessments.[115]

The 2010 law gave Colorado school districts the option of selecting the state's model evaluation system or developing their own systems by school year 2013-14. The law was amended in May 2014 to allow school districts to determine, for the 2014-15 school year only, the percentage of student academic growth, if any, to be factored into an educator's performance evaluation. Supporters of the revision cited the need for districts to analyze results from the state's new student assessments (Partnership for Assessment of Readiness for College and Careers, or PARCC, beginning in the 2014-15 school year) before they are used to evaluate teachers.[116]

Colorado has had the support of the U.S. Department of Education (USDOE) in making changes to its teacher evaluation system and other education reform efforts. In December 2011, the state was awarded a $17.9 million Race to the Top Phase 3 grant. One of the four key areas the grant was to be used for was to design the state's educator evaluation system, an effort that was already underway.[117] In early 2012, the USDOE granted Colorado a waiver under No Child Left Behind, allowing the state to use its own accountability model beginning in school year 2013-14, including the 2010 changes made to its teacher evaluation system, rather than the NCLB model.[118] In early July 2014, the USDOE granted Colorado a one-year waiver extension.[119]

1. Which teachers are required to have statistical growth model data as part of their evaluations?

Those teachers of grades and subjects for which state assessments are administered in consecutive years will receive CGM results: English language arts and mathematics grades 4 through 11, and grades 1 through 12 in English language proficiency.[120] Beginning in the 2014-15 school year, Partnership for Assessment of Readiness for College and Careers (PARCC) assessments for English Language Arts and Mathematics are scheduled to replace the Transitional Colorado Assessment Program (TCAP) tests.[121]

2. What weight is given to statistical growth model data in teacher evaluations?

In Colorado, beginning in 2015-16, 50 percent of a teacher's annual evaluation must be based on student academic growth. All Colorado teachers are required to have *measures of learning* (i.e., growth) scores as part of their evaluations, but only about one-quarter of teachers receive results that are generated through the Colorado Growth Model.[122]

For those teachers who receive them, CGM results must comprise some portion of the student academic growth component, but do not have to count for the full 50 percent. Each local school district must determine the weight to give the CGM scores within the growth component of the teacher evaluation. The Colorado Department of Education has imposed no minimum on the percentage that must be derived from CGM results.[123]

In 2014-15 only, the first year the PARCC tests are expected to be administered, districts may opt to have no student growth factors included in teacher evaluations, the result of a legislative revision signed into law in May 2014.[124] Unless further legislative changes are made, student growth will factor into teacher evaluations at the full 50 percent level in all districts beginning in 2015-16, but will not be tied solely to results from the Colorado Growth Model.

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

For teachers without CGM scores, the 50 percent portion of their evaluations related to student growth may include the same alternative elements as teachers with CGM scores: other state assessments, teacher-developed assessments, and student learning objectives. Local school districts determine which measures best reflect student growth and how the results from the selected student learning measures will be scaled for expected growth. Districts determine what constitutes results that are much less than expected, less than expected, expected, and more than expected based on where students or groups of students began in order to be included as a measure of student learning. The general standard is that students make at least a year's worth of growth in a year's time.[125]

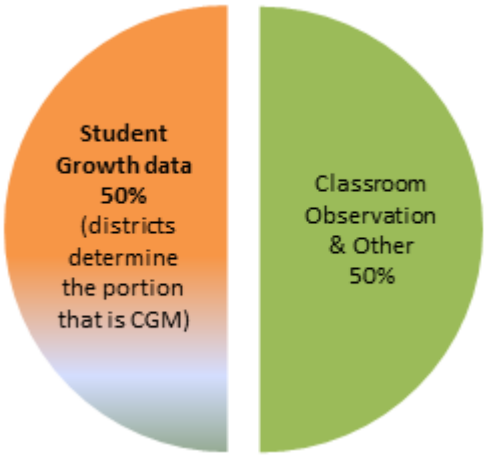
Using Race to the Top funds, the CDE created the Colorado Content Collaboratives, groups of Colorado educators who review and create assessments aligned to the state's academic standards and which may be used for educator evaluation purposes.[126]

4. What components are part of teacher evaluations in addition to student growth data?

In the Colorado State Model Educator Evaluation System, 50 percent of an educator's evaluation is based on instructional practices and 50 percent on multiple measures of student learning (i.e., student growth). The 50 percent of a teacher's evaluation comprising the instructional practice score is derived from the "Rubric for Evaluating Colorado Teachers" that outlines five quality standard areas: content, environment, instruction, reflection, and leadership. Each standard is associated with a number of elements that are scored individually. On the instructional practice scale, teachers are rated as exemplary, accomplished, proficient, partially proficient, or basic.[127]

5. How are teacher evaluation performance ratings applied to personnel decisions?

Exhibit 1: Colorado's Components of Teacher Evaluation, 2013-14



To develop a teacher's final effectiveness rating, the instructional practice scores and measures of student growth data are combined to determine an overall effectiveness rating of ineffective, partially effective, effective, or highly effective.[128] Teachers who receive ratings of effective or highly effective for three consecutive years will receive tenure (also referred to in Colorado as non-probationary status). Teachers who receive partially effective or ineffective ratings for two consecutive school years will be given probationary status and may be subject to non-renewal of their contract, though they may not necessarily lose employment.[129] Districts determine on an individual basis whether teachers who lose tenure will be re-employed.[130]

All teachers, regardless of status, are required to develop annual professional growth plans and new measures of student learning designed to address areas identified through the evaluation process and professional development or training needed.[131]

6. What type of statistical growth model is used for teacher evaluation?

Colorado has used the Colorado Growth Model (CGM) since 2009 to measure student achievement and growth on statewide assessments each year and student academic growth over time. The CGM is implemented using an open-source (i.e., available for free to any users) statistical software program.[132] Before its adoption as part of the teacher evaluation system, the Colorado Department of Education (CDE) had been using CGM results for school and district accountability, but CGM was not used at the teacher level.[133] The description provided here refers to the Colorado State Model Educator Evaluation System. Districts in Colorado, which exercise strong local control, have flexibility in deciding how to use the CGM for teacher evaluations.[134]

The CGM calculates growth using *student growth percentiles*, which describe how much growth each individual student makes relative to their academic peers. Student growth percentile scores range from 1 (lowest growth) to 99 (highest growth).[135] The CGM compares each student's current achievement to students in the same grade throughout the state who had similar state assessment scores in past years. The model then estimates a student growth percentile for each student. A student growth percentile of 60, for example, indicates that a student grew as well as or better than 60 percent of the student's peers.[136]

The CGM is also used to calculate a *median growth percentile*, which summarizes student growth rates by district, school, grade level, or other group of interest, such as a classroom. The median growth percentile is calculated by taking the individual student growth percentiles of all students in the

group being analyzed, ordering them from lowest to highest, and identifying the middle score (the median).[137]

The CGM can be used to attribute academic growth to a teacher by calculating the median student growth percentile of students linked to the teacher. Once conditional percentile ranks are computed for all students, teachers are assigned a score equal to the median conditional percentile rank of the students within their class. The scores can be used to form rankings of teachers by their estimated effectiveness.[138]

The CGM does not account for any student background characteristics, such as socioeconomic status, other than same-subject scores from previous years.[139] Colorado State Board of Education Rule 301.87 requires that student test results from two consecutive grades be used to generate CGM results associated with teacher evaluations.[140] The CGM includes additional years of student test scores when these earlier scores are available.[141] The model makes no statistical adjustments for teachers who teach small numbers of students. Districts using the model for teacher evaluation purposes must make several determinations that affect how growth is attributed to teachers; for example, districts must determine the minimum number of students in a classroom a teacher must have to generate an individual growth score, the minimum number of days of attendance for a student's score to count, and whether a reported growth score for a teacher represents only the current year score or a multi-year average.[142]

Currently, the Colorado Department of Education is not providing teacher level results through the Colorado Growth Model and is unsure when that will occur.[143] In 2009, the Colorado General Assembly created an advisory panel of state and national experts to make recommendations to the State Board of Education about the use of academic growth for accountability purposes.[144] In early 2014, the panel recommended against the use of median growth percentiles at the teacher level, but provided suggestions for districts that choose to do so: use at least two years of median growth percentiles, use confidence intervals, ensure that median growth percentiles are based on a sufficient number of students, and exercise caution in developing cut points for categories that describe teacher effectiveness.[145]

7. What issues and concerns does Colorado face from its use of statistical growth model data as part of teacher evaluations?

Colorado has not yet implemented the growth portion of its teacher evaluation system. Districts will be required to use teacher growth data beginning in the 2015-16 school year.[146]

DISTRICT OF COLUMBIA PUBLIC SCHOOLS[I]

Background

The District of Columbia's Teacher and Leader Evaluation Requirements were created in conjunction with the District's Race to the Top commitments and No Child Left Behind flexibility waiver, leaving many evaluation decisions to local districts. The Teacher and Leader Evaluation Requirements apply to District of Columbia Public Schools (DCPS) and all charter schools participating in Race to the Top, which together account for about 90 percent of the District's students.[147] DCPS' teacher evaluation system, IMPACT, was adopted in 2009-10 and aligns with the district's teacher evaluation requirements.

The U.S. Department of Education granted DCPS and other area education agencies a No Child Left Behind flexibility waiver in 2012, and a one-year waiver extension for 2014-15. DCPS announced in June 2014 that student growth model data would not be used in teacher evaluations during 2014-15 due to the transition to its new testing program, ahead of the federal policy change allowing states to delay implementation of the NCLB waiver requirement to use student test data in teachers' evaluations.[148]

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

Math teachers in grades 4-8 and English language arts teachers in grades 4-10 have had growth model data based on the District of Columbia's Comprehensive Assessment System (DC CAS) test used for their evaluations. In 2013-14, about 15 percent of teachers had individual value-added data.[149] DCPS plans to implement standardized assessments for more subjects and grades in future years which will increase the number of teachers with growth model scores.

2. What weight is given to statistical growth model data in teacher evaluations?

Growth model scores are weighted as 35 percent of teacher evaluations.[150] The original IMPACT system was revised in 2012-13, based on teacher feedback, reducing the weighting for growth model scores from 50 percent to 35 percent.[151] DCPS announced in June 2014 that it would not use student growth model scores in 2014-15 because of the district's transition to PARCC tests.[152] DCPS will incorporate value-added data into teacher evaluations again during 2015-16.[153]

[I] This profile focuses primarily on District of Columbia Public Schools (DCPS), one of the local districts under the oversight of the Office of the State Superintendent of Education, which functions in ways similar to a state education agency. District of Columbia has 58 local education agencies (LEAs) including DCPS and each charter school operator.

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

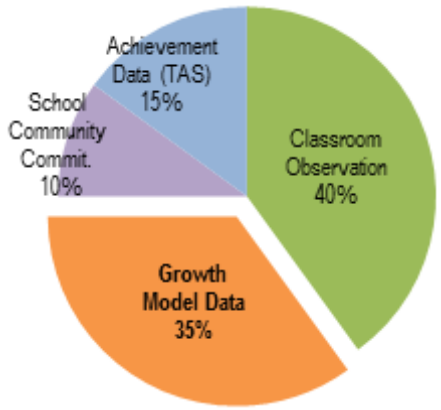
Instead of developing alternatives to student growth data for teacher in non-tested subjects and grades, DCPS weights the classroom observation component more heavily: 75 percent instead of the 40 percent used for teachers with growth model data.[154] The district has not indicated plans to develop alternative measures.[155]

4. What other components are part of teacher evaluations in addition to student growth data?

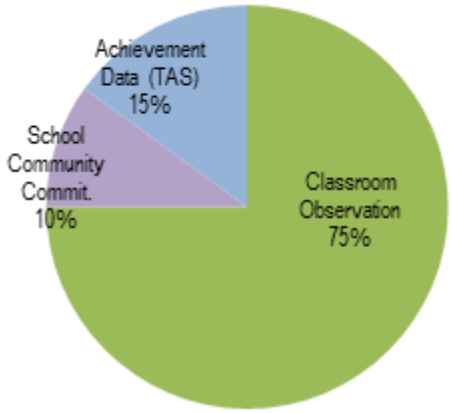
In addition to the 35 percent for growth model scores, 40 percent of the total evaluation score is based on classroom observations, 15 percent on teacher-assessed student achievement data (TAS), and 10 percent on commitment to the school community. The evaluation's core professionalism component does not add to evaluation scores but can cause score deductions.[156]

Exhibit 1: District of Columbia's Components of Teacher Evaluation, 2013-14

Teachers with individual growth model data



Teachers without individual growth model data



Note: For both pie charts, up to 40 points (on a 400-point scale) can be deducted for failure to meet core professionalism requirements.

The combination of growth model scores and teacher-assessed student achievement data (TAS) means that 50 percent of the evaluation is based on student performance measures. The TAS component is similar to student learning objectives used in other evaluation systems.[157] The classroom observation component is based on the Teaching and Learning Framework, which covers the areas of planning, teaching, and increasing effectiveness. Only the teaching portion, with nine key activities for classroom performance, has been used for evaluation purposes. Each component has a scoring rubric.

Commitment to school and community is based on school administrators' assessment of teachers' support of school initiatives, support of special education and English language learner programs at the school, promotion of high academic and behavioral expectations, partnership with students' families, and instructional collaboration with colleagues.

Teachers must meet the four basic requirements within the “Core Professionalism” component, as assessed by school administrators according to a rubric twice a year, in order not to lose points on their evaluation scores. The requirements are:

- no unexcused absences
- no unexcused late arrivals
- following school and district policies and procedures, and
- respectful interaction with stakeholders.

5. How are teacher evaluation performance ratings applied to personnel decisions?

Teachers earn one of five rankings based on their evaluations: highly effective, effective, developing, minimally effective, and ineffective.

The DCPS teacher compensation system is aligned with its evaluation system, allowing teachers earning rankings of “highly effective” to earn annual bonuses and eligibility for the career path program, which offers opportunities to accelerate increases in base pay for those who work in high-poverty schools.[158] Larger bonuses are available for highly effective teachers who teach in grades and subjects with growth model scores. Teachers earning effective rankings are also eligible for the career path program.

Teachers ranked “developing” are not eligible for pay increases or career path advancement, but do get priority for professional development; however, if no improvement is shown after three years, teachers ranked as developing are subject to dismissal. Teachers ranked “minimally effective” are given priority for professional development and subject to dismissal if they do not improve after two years. Teachers ranked “ineffective” are subject to dismissal.

DCPS first used teacher evaluation ratings for high-stakes personnel decisions in 2010, when 241 teachers were fired, most for poor performance on their IMPACT evaluations; the district has continued to fire its lowest performing teachers each year.[159] The percentage of teachers rated highly effective has grown from 16 percent in 2010-11 to 29 percent in 2012-13, while the percentage of ineffective and minimally effective teachers has declined from 15 percent to 6 percent over the same period.[160]

6. What type of statistical growth model is used for teacher evaluation?

DCPS has contracted with Mathematica Policy Research to calculate growth model scores for teachers. A teacher’s value-added is the difference of the classroom average of students’ actual scores from the classroom average of their expected scores.[161] Each teacher’s value-added results are compared to the value-added results of other teachers district-wide. The model includes a procedure to adjust teachers’ estimates within each grade level to reflect the assumption that the variation in teacher effectiveness is similar across grade levels.[162]

Students' expected scores on the DCPS standardized test are estimated based on their test scores from the past year and those of other students who received similar scores in the past year.[163] An auxiliary year of data is also included in the model.[164]

The model also controls for the student characteristics of poverty (eligibility for free or reduced-price lunches), special education status, and limited English proficiency status, as well as for time in the classroom based on previous year attendance and whether the student transferred schools mid-year. The model controls for peer effects (how classmates may influence achievement) by factoring in a class's average test score from the previous year and the variation in classmates' scores from the previous year.[m]

Teachers' current year value-added effect score is used for evaluations.[165] Standard errors for the scores are not included. Teachers must have at least 15 students with test scores from the previous and current year in order to receive a growth model score.[166] Students must be enrolled in a class for a minimum of five weeks for their scores to be included in value-added calculations. Attendance for students who meet the five week minimum is controlled for as a student characteristic.

Students who have missing test score data are not included in teachers' value-added calculations, but if they have test score data and are missing other data, such as characteristics used for control factors, their data is inferred from similar students.[167] Mathematica uses a shrinkage method to adjust for small samples and standard error.

7. What are the issues and concerns facing District of Columbia from its use of statistical growth model data as part of teacher evaluations?

In 2012-13, approximately 40 teachers in DCPS received incorrect evaluations because of errors in the way growth model scores were calculated. One teacher was fired as a result but later reinstated when the errors were discovered.[168]

DCPS was one of the earliest districts to link high-stakes personnel decisions, like bonus pay and dismissals, to evaluation scores. A 2011 test cheating scandal received media attention as a possible consequence of such linkages.[169] Subsequent investigations concluded that, although some anomalies had been identified at one elementary school, cheating was not widespread and improvements could be made in test security procedures.[170]

[m] The theory behind these controls for peer effects is that (a) classmates that are higher-achieving, on average, can push other students to higher learning gains and, conversely, classmates that are lower-achieving, on average, can pull other students to lower learning gains, and (b) the wider the variation of student achievement levels in a classroom, the more difficult it is for a teacher to ensure that all students are achieving learning gains. Eric Isenberg and Elias Walsh, *Measuring Teacher Value Added in DC, 2012–2013 School Year: Final Report*, Mathematica Policy Research, Sept. 11, 2013, p. 2, http://dcps.dc.gov/ (accessed June 18, 2014).

FLORIDA

Background

In 2011, the Florida Legislature passed the Student Success Act, which revised the state's requirements for districts' teacher evaluation systems, and required the Commissioner of Education to approve models for measuring student growth.[171] The 2011 law was adopted as part of the state's efforts related to its Race to the Top grant. Florida's NCLB waiver has been extended through 2014-15, and the state has been notified that it may be eligible for a longer renewal. The state has been phasing in its new teacher evaluation system over the past several years. The full implementation at the local district level is scheduled for 2015-16.[172]

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

Florida requires state-provided value-added data to be used in districts' evaluations of teachers in English language arts (grades 4-10), math (grades 4-8), and Algebra I. Formerly based on the Florida Comprehensive Assessment Test (FCAT), growth model data from end-of-grade exams will be based on the Florida Standards Assessment beginning in 2014-15. Algebra I results will continue to be based on the end-of-course assessment. The Florida Department of Education expects to approve growth models for additional statewide assessments in the future.

2. What weight is given to statistical growth model data in teacher evaluations?

Teachers with state-provided value-added data are required to have it weighted as a minimum of 50 percent of their evaluations.[173] If less than three years of individual value-added teacher data are available, the student growth component may be reduced to a minimum of 40 percent of the teacher's evaluation.[174]

For grades and subjects with statewide assessments and a state-approved student learning growth formula, school districts must use the state-provided value-added data for the student growth component of teachers' evaluations. The Florida Department of Education provides districts the value-added data in several formats from which local districts can choose for their evaluation systems. (Formats include a three-year aggregate, one-year aggregate (district computes a three-year combination), separate grade and subject reports (district computes a three-year combination), or student-level reports of percent meeting or exceeding their predicted scores.)[175] Districts also set the cut scores for how value-added data is converted into overall teacher performance ratings. The Board of Education is expected to set state requirements for value-added performance ratings in 2015, which may reduce districts' flexibility in how they incorporate value-added in teacher evaluations.[176]

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

Local school districts determine the student assessments and growth measures used for calculating the student growth component for teachers without state-provided value-added data. The district growth measures must also be weighted at a minimum of 50 percent of the overall evaluation score and can be based on:

- statewide assessments (without a state-approved growth formula),
- nationally recognized standardized assessments,
- industry certification assessments,
- district-developed or district-selected end-of-course assessments, and
- educator-selected assessments.[177]

Some growth measures that districts can choose include:

- proficiency models,
- simple growth models,
- advanced statistical models, and
- student learning objectives.

Local districts may contract out the calculation of growth data or calculate it themselves for subjects and grades not covered by state-provided value-added. Districts may request permission to use certain other data for teachers of courses without state assessments, such as student achievement data rather than, or in combination with, student growth data or state-provided student growth data for students who are enrolled in a course that is not directly tied to the state assessments but is in an associated subject area.[178] An interim provision, expiring at the end of July 2015, allows districts to use measurable learning targets on local district assessments or to use state-provided student growth data for an instructional team's students for all teachers assigned to an instructional team.

4. What other measures are part of teacher evaluations in addition to student growth data?

State-required components of districts' teacher evaluation systems include instructional practice, evaluated through a classroom observation rubric, and professional and job responsibilities as adopted by the State Board of Education, plus any adopted by a local school board.[179] Each district determines the proportion that each component is weighted in the overall teacher evaluation, as long as the student performance component is at least 50 percent.[180]

Most Florida districts use one of three instructional practice rubrics for classroom observation: the Marzano, Danielson, or Educational Management Consultant Services (EMCS) model.[181] All rubrics must align with the Florida Educator Accomplished Practices, six areas of effective teaching defined in State Board rules.[182] The professional and job responsibilities component is defined as duties outlined in the Code of Ethics and the Principles of Professional Conduct, which together address respect, non-discrimination, fairness, truthfulness, and other qualities expected of teachers.[183] Districts may include

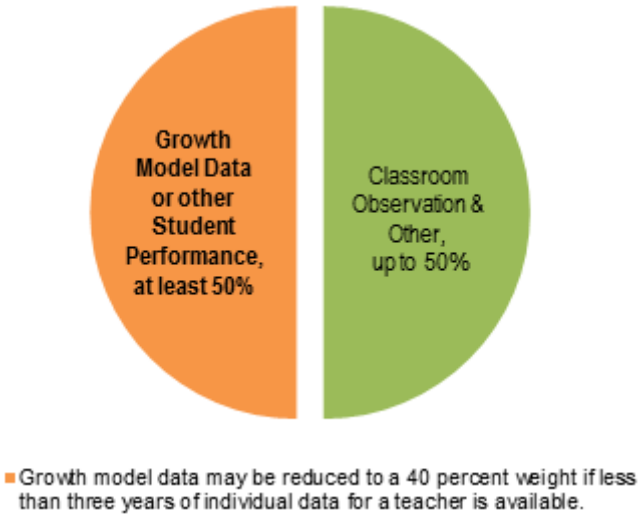
additional duties within this component and may include measurement of this component within the instructional practice component.

5. How are teacher evaluation performance ratings applied to personnel decisions?

Florida requires districts to rate teacher performance using four categories:

- highly effective,
- effective,
- needs improvement, or – for teachers in their first three years of teaching whose performance is in this category – developing, and
- unsatisfactory.

Exhibit 1: Florida Components of Teacher Evaluation, 2013-14



State law requires the State Board to adopt rules on performance levels that will result in a teacher being rated unsatisfactory if student learning standards are not met.[184] Similarly, students must achieve certain learning standards for a teacher to receive effective and highly effective ratings. Standards for each teacher performance level are required to be set by the State Board beginning in 2015-16.

Part of Florida's Student Success Act required that districts adopt new performance pay plans that

- provide annual raises only to teachers who earn a rating of effective or highly effective, unless they are "grandfathered" in under continuing contracts or professional service contracts, defined in previous tenure laws, or for which the school district has not yet adopted student growth measures,
- require highly effective teachers on performance pay plans to earn a larger raise than that available to any other teacher of the same classification, whether they are grandfathered in or not, and
- permit teachers grandfathered in under continuing contracts or professional services contracts to opt into the performance salary schedule if they switch to annual contracts.[185]

All Florida teachers on annual contract as of, or hired after, July 2014 are placed on districts' performance salary schedules. Beginning in 2014-15, teachers on their districts' performance salary schedules will qualify for the next year's salary increases based on their district's plan and their evaluation results.[186]

The Student Success Act also effectively put an end to tenure except for teachers who already had it, requiring that all teachers hired after July 2011 work under annual contracts.[187] Annual contracts may not be awarded to or renewed for teachers who have received two consecutive performance evaluation ratings of unsatisfactory, two ratings of unsatisfactory within a three-year period, or three consecutive

ratings of needs improvement or unsatisfactory in any combination. Teachers with annual contracts may be dismissed for just cause, which includes the same poor evaluation rating criteria as those that prevent awarding of annual contracts, and teachers with continuing contracts can be returned to annual contract status.[188]

Florida's overall teacher evaluation results for 2012-13 show that most Florida teachers were rated highly effective (32 percent) or effective (66 percent).[189] About 2 percent of teachers received rankings of needs improvement or developing in 2012-13 and less than half of 1 percent were rated unsatisfactory. Preliminary results for 2013-14 show very similar numbers, with 97 percent of teachers rated highly effective (42 percent) or effective (55 percent), 2 percent rated needs improvement or developing, and less than half of 1 percent rated unsatisfactory.[190] Teachers with the highest ratings may be subject to fewer classroom observations the following year. Teachers rated "needs improvement" or "developing" may receive more classroom observations, professional development, or be assigned to work with a more experienced teacher as determined by the district. State law prescribes that:

- districts' primary consideration for employee promotion must be their effectiveness under the evaluation system,
- layoffs due to reduction of positions must be based on educational program needs and employee performance evaluations, and
- teachers who supervise student teachers must have earned a prior year rating of effective or highly effective.[191]

6. What type of statistical growth model is used for teacher evaluation?

Florida contracted with the American Institutes for Research (AIR) in 2011 to help develop and implement its value-added model. Individual teacher value-added data are reported as the average student growth of a teacher's students above or below expected learning growth of similar students in the state, based on the student, classroom, and school factors controlled for in the model.[192] The value-added estimates are reported for teachers as one-, two-, and three-year averages, except for Algebra I EOC results, which are only reported as single-year scores.[193]

The model establishes an expected growth for each student, based on growth for similar students in the same grade in the same year. In the AIR model, students' previous test scores are controlled for directly as a specific predictor variable.[194] The model uses up to two years of prior test data, but only the immediate prior year's test score is required for the student to be included.[195] The model controls for other student and classroom factors for both the end-of-grade test and the EOC test results as follows:

<u>End-of-grade test</u>[196]

Student Characteristics

- past assessment performance - up to two prior test scores,
- students with disabilities (14 indicators),

- English language learners (four indicators),
- gifted status,
- attendance (percent of days present),
- mobility (number of transitions, e.g., from one school or district to another),
- difference from typical age in the student's grade (retention indicator), and
- number of subject-relevant courses.

Classroom characteristics

- class size, and
- similarity among students' past year test scores.

End-of-course test[197]

Student characteristics

- past assessment performance – up to two prior year-to-year math scores,
- students with disabilities,
- English language learners,
- gifted status,
- attendance,
- mobility, and
- difference from typical age in student's grade.

Classroom characteristics

- class size,
- variance among students' past year test scores,
- percent of gifted students in class,
- percent of students at typical age for the grade, and
- average prior test score for class.

State law prohibits the statistical growth model from setting different expectations of student growth based on gender, race, ethnicity, or socioeconomic status.[198]

Because school-wide factors, representing school leadership, culture, environment, and other impacts on student learning are considered partially attributable to individual teachers, half of the school component is added into teachers' individual value-added estimates.[199] Note that this school component, which accounts for about one-third of a teacher's total value-added estimate, is different than the school's overall value-added estimate; it is the school effect attributed to teachers apportioned by the number of students they teach and is different for each teacher.[200]

Value-added data is estimated for each grade and subject. Students are not included in the end-of-grade value-added models if they do not have the previous year's test score in the same subject and are not included in the EOC model if they do not have the previous year's math score.[201] Students' attendance is incorporated into the model as a percentage of days attended divided by days absent;

there is no minimum attendance requirement.[202] The value-added reports to districts include standard error so that "districts can construct the confidence interval needed for performance classification."[203] Teachers with at least two students in their classroom receive state value-added scores.[204] Districts determine which value-added data to use and how to convert them into teacher evaluation rankings. (See more at Question 2.)

7. What are the issues and concerns facing Florida from its use of statistical growth model data as part of teacher evaluations?

Two lawsuits have been filed in Florida related to the use of statistical growth models in teacher evaluation. In the first, a newspaper filed suit against the state's Department of Education for its failure to provide teachers' value-added scores requested under open public records laws.[205] Florida law requires teachers' overall evaluation ratings to remain confidential for one school year.[206] The legal question was whether the value-added data, as one element of evaluations, was subject to disclosure or could similarly remain confidential for one year. A Florida Court of Appeal overturned an earlier circuit court decision, ruling that data used to prepare an evaluation does not have the same protection as the full teacher's evaluation.[207] Based on the court ruling, the Department of Education released teachers' value-added scores for 2011-12 and 2012-13 in February 2014.

The second lawsuit, filed in 2013 by seven teachers and their unions, argued that the state's teacher evaluation system violated their constitutional rights for equal protection and due process by evaluating the teachers partly on the basis of test scores of students and in subjects they had not taught, claiming such policies were irrational and arbitrary.[208] The U.S. district judge ruled that despite "the unfairness of the evaluation system as implemented," the case rested on whether the evaluation policies are "rational within the meaning of the law."[209] The judge ruled against the teachers' claims that the evaluation system was unconstitutional. State law was revised to clarify that teachers' evaluations must be based on performance data of their students.[210]

LOS ANGELES UNIFIED SCHOOL DISTRICT

Background

California state law, as interpreted by the decision in a 2013 lawsuit involving the Los Angeles Unified School District (LAUSD), requires that teacher evaluations must incorporate measures of student progress, but does not specify methods of measurement.[211] In 2009-10, LAUSD began piloting a school level value-added model. In 2012-13, the district extended the pilot to include teacher level value-added analysis on a "no-stakes" basis (i.e., the results could have no negative repercussions for teachers).[212]

For the last few years, LAUSD has been developing a new teacher evaluation system, aided partly by a $49.9 million Teacher Incentive Fund grant from the U.S. Department of Education awarded in September 2012. The grant requires that the district use a teacher evaluation system that includes both observations and valid measures of student growth.[213] LAUSD also received a No Child Left Behind waiver in late 2013, as part of an eight-district consortium. One of the waiver's requirements is to use a student growth model for teacher evaluation.[214]

The use of student growth as a part of the teacher evaluation process continues to be the subject of significant union pushback for LAUSD.[215] The current teacher evaluation process in LAUSD is governed by a collective bargaining agreement between United Teachers Los Angeles (UTLA) and the district, as well as a supplemental agreement developed as a condition of the 2013 lawsuit.[216] Teacher evaluation is a mandatory subject of collective bargaining. Teacher evaluation processes are, therefore, locally negotiated.[217]

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

Teacher level value-added data, termed Academic Growth over Time (AGT) in LAUSD, was provided in 2012-13 for teachers of:[218]

- grades 3-9, English language arts
- grades 3-8, mathematics
- Algebra I
- Algebra II
- Geometry
- Biology
- Chemistry
- Physics
- Integrated Science I
- Science Grade 8
- US History
- World History

Value-added data (AGT data) was provided for about 55 percent of the district's teachers (i.e., those that taught subjects and grades that had associated state tests) in 2012-13.[219]

Because of a 2013 state law, LAUSD is currently unable to calculate value-added estimates for teachers. California Assembly Bill 484 suspended most of the state's required tests while the Smarter Balanced ELA and math assessments were field-tested and phased in during the 2014-15 school year. The suspension includes all high school end-of-course tests, many of which are used to calculate teacher level value-added estimates. LAUSD staff are uncertain of the future of the state's assessments, other than the Smarter Balanced tests, and how this might affect the calculation of value-added estimates.[220]

2. What weight is given to statistical growth model data in teacher evaluations?

Although LAUSD officials have been working toward including a weight limited to 30 percent of a teacher's total evaluation score, the district currently does not use any of the growth model estimates in the teacher evaluation process. Value-added scores were last supplied at the teacher level in 2012-13, and an LAUSD official indicates that it may be 2016 before the district is able to calculate value-added estimates for teachers again.[221]

Any use of a value-added estimate in teacher evaluations will have to be negotiated with the teachers' union before it can be implemented. The 2013 supplemental agreement that governs the district's current teacher evaluation process (along with the UTLA collective bargaining agreement) states that

> Individual AGT scores . . . are to be used solely to give perspective and to assist in reviewing the past CST [California State Test] results of the teacher, and shall neither form the basis for any performance objectives/strategies nor be used in the final evaluation.[222]

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

For teachers without state-assessment data that can be analyzed using the value-added model, additional data sources may be used, including:

- school level results;
- school level based goals for pupil progress and achievement, such as attendance rates, suspension rates, ELL progress and reclassification rates, graduation and dropout rates, AP course enrollment and passage rates, and others;
- periodic/benchmark assessment data, where available and appropriate to the students and curriculum, such as Periodic Assessments (elementary) and DIBELS, and other standards-based assessment data/student work samples, projects, portfolios;
- assessment data that documents pupil performance, such as an Independent Reading Level assessment, Developmental Reading Assessment, Qualitative Reading Inventory;
- pre- and post-assessment data, such as the start and culmination of a semester or other unit of study;

- curriculum-based examinations and similar culminating activities; and
- for special education students with IEPs, various diagnostic assessments to measure progress toward previously-identified goals.[223]

4. What components are part of teacher evaluations in addition to student growth data?

LAUSD began creating a new teacher evaluation system after a 2009-10 task force made recommendations to the Board of Education and, in the 2014-15 school year, is continuing to develop the evaluation process. The framework for the district's new evaluation system, the Educator Growth and Development Cycle, consists of five standards: planning and preparation; classroom environment; delivery of instruction; additional professional responsibilities; and professional growth. Rubrics provide descriptions of the levels of performance for each of the elements within the framework.[224]

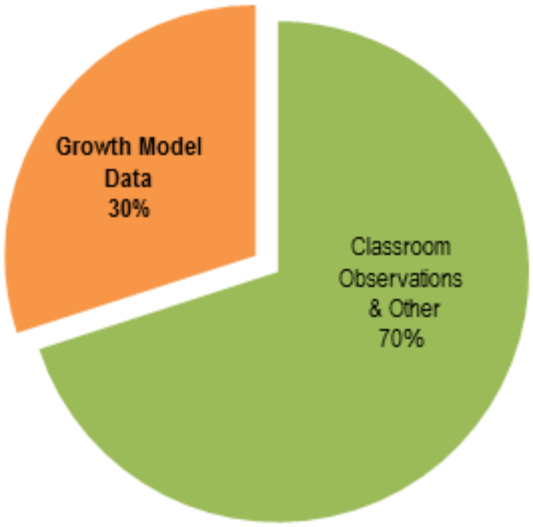
The district describes its new evaluation process as a system of "multiple measure reviews with aligned professional development and support structures." Performance measures are based on observation of teacher practice, contributions to student learning outcomes, stakeholder feedback, and contributions to school community.[225]

Under full implementation of LAUSD's evaluation system, the district plans that teachers' final ratings will be ineffective, developing, effective, or highly effective.[226] Under the evaluation system currently in use, which is governed by the collective bargaining agreement, teachers receive one of two ratings on the final evaluation: meeting standard performance or below standard performance.[227]

5. How are teacher evaluation performance ratings applied to personnel decisions?

A teacher who receives an overall below standard rating may be subject to an administrative review to determine the need for disciplinary procedures, governed by the UTLA collective bargaining agreement.[228] A "below standard performance" rating is subject to grievance procedures.[229]

Exhibit 1: LAUSD's Components of Teacher Evaluation – PROPOSED



The TIF grant application for LAUSD indicates that the district's intent is to use the overall teacher evaluation ratings as "the most significant, in determining which educators are eligible for tenure approval, retention incentives, career ladder opportunities, and additional compensation." The 2012 application also states that the district would be working toward developing a prescribed weight for each measure of the evaluation system through negotiations with the teachers' union "to ensure [the student growth measure] is significantly weighted in the formula."[230] According to an LAUSD official, these negotiations are ongoing as of fall 2014.[231]

Under the union agreement with LAUSD, the assessment of student progress will not be the “sole, primary or controlling” factor in the final rating determination, but is to be considered “an important, but clearly limited, part of the overall evaluation of the employee’s performance.”[232]

6. What type of growth model is used for teacher evaluation?

In partnership with the Value-Added Research Center (VARC) at the University of Wisconsin-Madison, LAUSD has developed an Academic Growth Over Time (AGT) model that uses a value-added methodology. The LAUSD AGT model uses students’ standardized test scores from prior years combined with student demographic data to create individual student growth predictions. Students’ predicted results are compared to their actual assessment results in order to calculate a value-added estimate, which may be attributed to the teacher. The model controls for factors that may affect student learning and have an effect on test scores, including prior test scores, grade level, gender, ethnicity, free/reduced price lunch status, English language learning status, special education status, continuous enrollment, and homelessness.[233] The model does not control for class size or for student mobility.[234]

In cases where students do not have subject-specific prior test scores, the model uses students’ test scores from multiple subjects to predict performance. For example, grade 5 science does not have a grade 4 science assessment available to measure prior science knowledge of students. Instead, each student’s grade 4 math and grade 4 English language arts tests are used to predict their grade 5 science performance. The model analyzes the grade 5 science performance of demographically similar students with the same grade 4 math and English language arts scores to calculate predicted outcomes for students.[235]

At the teacher level, a value-added score is reported only if a teacher had a sufficient number of students (i.e., 11 or more). Students are only included in the analysis if they were continuously enrolled in the same school from a specific date in October through the date in May when tests are conducted, have prior year test data, and have a complete demographic profile (i.e., ethnicity, gender, free or reduced price lunch status, special education status, homelessness, and English language learner status).[236]

The model employs a shrinkage method to reduce the possibility of incorrectly categorizing teachers as either effective or ineffective, especially if they have small numbers of students.[237]

In reports, AGT results are presented as point estimates surrounded by a 95 percent confidence range. Teachers receive single-year AGT measures, multiple-year AGT measures that reflect student growth over as many as three years, and single-year results for student subgroups within the teacher’s classroom. If a teacher has not been teaching long enough to have all three AGT measures, the average score is based off the maximum amount of data. A first-year teacher’s average and single-year results will be the same.[238]

AGT methodology can be applied to student data from both norm-referenced and criterion referenced tests. The actual AGT estimates for schools and teachers are norm-referenced, comparing the academic growth of students in a particular school or classroom to the growth of similar students across the district. For each grade level and time period, AGT compares a school or teacher to the rest of LAUSD for that same grade level and time period.[239]

7. What issues and concerns does LAUSD face in using statistical growth model data as part of teacher evaluations?

Conflict with the teachers' union, particularly concerning the use of student growth data in teacher evaluations, is a significant concern for LAUSD.

Another concern is the public disclosure of teacher value-added data, which LAUSD has resisted and which was the subject of a 2012 lawsuit filed by the *Los Angeles Times* against the district. The newspaper requested the value-added data, which the district supplied with teachers' names redacted. The newspaper sued for release of teachers' names with the data. An August 2013 court decision (Los Angeles Times Communications LLC v. Los Angeles Unified School District) allowed the newspaper to publish performance ratings of individual teachers. LAUSD appealed, and a 2014 Court of Appeals ruled that teacher names are exempt from disclosure in response to a public records request for the district's scores for teacher effectiveness. One facet of the case has yet to be resolved: whether "location codes," which would allow the newspaper to link the redacted AGT scores to the schools where teachers are located, can be publicly released. The Court of Appeal returned this matter for resolution to the trial court.[240]

NORTH CAROLINA

Background

North Carolina began developing a new approach to teacher evaluation in 2007, in partnership with a private nonprofit organization, Mid-continent Research for Education and Learning (McREL). The state and McREL developed the North Carolina Educator Evaluation System (NCEES), used for the first time statewide in the 2010-11 school year. NCEES was based on the state's five professional teaching standards, which were adopted in 2007-08. In July 2011, as part of the state's successful Race to the Top effort, the North Carolina Department of Public Instruction added a sixth standard concerning teachers' contributions to student academic success, to be measured using a value-added metric.[241]

On May 29, 2012, the U.S. Department of Education (USDOE) approved North Carolina's application for a waiver from No Child Left Behind, which included a requirement to establish a teacher evaluation system that incorporates student achievement.[242] The USDOE has granted the state a one-year extension to its waiver through the end of school year 2014-15, indicating that if it remained on track to implement its planned teacher evaluation system, it would be eligible to apply for a longer renewal period in spring 2015.[243]

In February 2014, North Carolina was granted approval by the U.S. Department of Education to delay using its teacher evaluation system to inform personnel decisions until 2016-17.[244]

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

North Carolina calculates student growth values using results from the following state tests, which are part of the North Carolina Testing Program:[245]

- End-of-grade and end-of-course assessments for grades 3-8 English Language Arts and Mathematics, grades 5 and 8 science, and Biology, Math I, and English II
- Career and Technical Education State Assessments for Career and Technical Education courses at the middle school and high school levels
- NC Final Exams (formerly called NC Common Exams) for courses and subjects in grades 4-12 English Language Arts, Mathematics, Science, and Social Studies when an end-of-grade or end-of-course assessment is not administered

About 40 percent of North Carolina's public school teachers administer one of these exams and receive individual growth model data through the Education Value-Added Assessment System (EVAAS) by SAS, a statistical model originally developed by Dr. William Sanders for Tennessee.[246]

2. What weight is given to statistical growth model data in teacher evaluations?

North Carolina uses a "threshold" approach to its teacher evaluation system: each component is weighted equally, including the growth component (Standard 6). Teachers must meet a minimum level of performance on *each* component to be rated as effective.[247] For those teachers who receive a rating through EVAAS, the rating constitutes the entirety of their growth score (for Standard 6).[248] (See Question 4 for more about Standards 1 through 5.)

3. What student growth measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

With the assistance of 800 teachers, North Carolina created and adopted more than 50 new assessments for subjects and grades not covered by state tests, including science, social studies, high school courses, music, art, and physical education.[249] The North Carolina Department of Public Instruction has also developed a guide that outlines methods for determining the individual growth value for teachers without EVAAS data. The four methods are:[250]

- Analysis of Student Work (ASW): Teachers identify standards and collect evidence of learning. Reviewers examine evidence and determine if students have made growth.
    - Beginning in 2014-15, ASW is the growth measure used for teachers of Advanced Placement, Arts Education, Healthful Living, International Baccalaureate, and World Languages.[251]

- Pre-Post Test Growth Model: Teachers administer a pre-assessment at the beginning of the course or grade and a post-assessment at the end of the year or semester. The scores from the two tests are then analyzed to produce a growth score for the teacher.
    - Used for courses and grades with statewide assessments, but where EVAAS is not used; for example, early grades, which lack sufficient prior student test data.

- Specialized Areas: The district determines local ways of measuring growth.
    - Includes most teachers of Academically or Intellectually Gifted (AIG) and Exceptional Children (EC).

- Local Flexibility: In 2013-14 and 2014-15, districts and charter schools may use school-wide growth data to determine growth ratings for specific groups of educators.

4. What components are part of teacher evaluations in addition to student growth data?

The North Carolina Board of Education requires principals to evaluate teachers using six professional teaching standards. Five of the standards focus on teacher performance. Four of these are largely determined through classroom observations and completion of a rubric by the observing principal; for the fifth standard, teachers complete a self-assessment using the same rubric.[252] (See Question 2 about the sixth standard, which measures how teachers contribute to student academic success.)

Exhibit 1: North Carolina Teaching Standards 1 – 5 and Performance Ratings

Teaching Standards 1 – 5	Performance Ratings
1. Teachers demonstrate leadership	▪ Distinguished ▪ Accomplished ▪ Proficient ▪ Developing ▪ Not Demonstrated
2. Teachers establish a respectful environment	
3. Teachers know the content they teach	
4. Teachers facilitate learning for their students	
5. Teachers reflect on their practice	

Source: Public Schools of North Carolina, North Carolina Teacher Evaluation Process, originally published 2009, revised 2012, p. 6, http://www.ncpublicschools.org/.

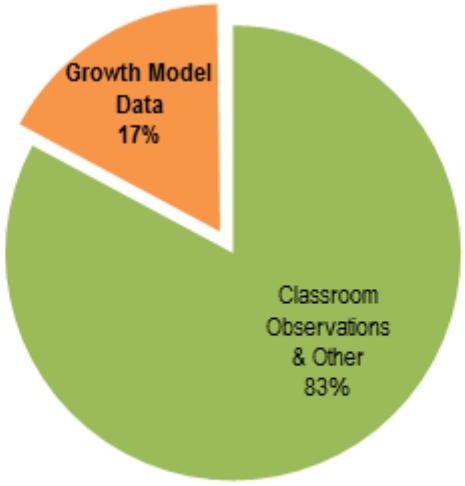
Each standard is weighted equally; teachers must meet a minimum level of performance on each to be considered effective.

5. How are teacher evaluation performance ratings applied to personnel decisions?

Under the evaluation system, teachers receive an overall effectiveness status in one of three categories:

- highly effective,
- effective, or
- in need of improvement.

Exhibit 2: North Carolina's Components of Teacher Evaluation



North Carolina's teacher evaluation process applies equal weight to all six standards. OREA extrapolated information about the state's process to create the pie chart.

These overall evaluation levels incorporate ratings from all six standards:[253] the five teaching performance standards assessed on a five-tier scale (see Question 4) and the sixth standard based on student growth, assessed on a three-tier scale (see Question 6). The criteria for combining the ratings into an overall status are shown in Exhibit 3.

Exhibit 3: North Carolina Teacher Effectiveness Levels and Criteria

Overall Effectiveness Status	Criteria
Highly Effective	Standards 1-5: Rated at least accomplished AND Standard 6: Exceeds expected growth
Effective	Standards 1-5: Rated at least proficient AND Standard 6: Meets expected growth
In Need of Improvement	Standards 1-5: Rated developing or below on any standard OR Standard 6: Does not meet expected growth

Source: North Carolina State Board of Education, Policy Manual, Policy ID Number TCP-C-006, Current Policy Data: Oct. 3, 2013, pp. 8-9, http://sbepolicy.dpi.state.nc.us/ (accessed July 25, 2014).

Teachers receive ratings annually on the first five standards (teaching performance) but the overall status ranking is not determined until after three consecutive years of student growth data have been collected, allowing the sixth standard (growth) results to be determined.[254] The U.S. Department of Education, as part of the state's amendment to its Race to the Top plan and waiver from NCLB, approved a timeline in which 2012-13 was the first of the three years of required data; 2014-15 will be the third year of data and the first year for which individual teachers' overall effectiveness rankings will be provided.[255]

Districts may choose whether to use the overall teacher evaluation status rankings for personnel decisions in school year 2015-16; they will be required to use them for personnel decisions beginning in school year 2016-17.[256] Local boards must use the performance standards and criteria adopted by the State Board, and may also adopt additional evaluation criteria and standards.[257]

According to an October 10, 2013, memo from the state to superintendents, written in anticipation of the state receiving USDOE approval to delay consequences for teachers based on their evaluations:[258]

- There will be no state-mandated consequences for a status of "in need of improvement" awarded at the beginning of the 2015-16 school year.
- The SBE will approve policies that require consequences for educators who are awarded "in need of improvement" status at the beginning of the 2016-17 school year.

Other personnel decisions based on evaluations include the type of individual growth plans for teachers, which may be directed, monitored, or mandatory depending on their performance and growth ratings, and subsequent dismissal, demotion, nonrenewal, or, in some cases, license revocation if required improvement does not occur.[259]

The state is moving toward compensating teachers based in part on their performance. The Governor of North Carolina created the Career Pathways for Teachers Fund, to help the state's school districts "transition to a rigorous, locally designed performance-based teacher compensation system." In the 2014-15 school year, districts are to develop and submit applications for funds, to be approved by the State Board. Select districts will pilot the model beginning in 2015-16. "Pathways must include highly-paid advanced roles for teachers consistently rated highly effective." By 2017-18, all districts must adopt a career pathways plan or use the default state plan.[260]

6. What type of statistical growth model is used for teacher evaluation?

In February 2012, the Board approved SAS Institute's Education Value-Added Assessment System (EVAAS), as the method for calculating growth scores for teachers.[261] North Carolina had been using EVAAS since 2007-08 for district and school level reports to inform educators' professional development needs,[262] but had not used the results for teacher evaluations.

In general, EVAAS uses a longitudinally merged database of student assessment scores to provide estimates of the progress of groups of students, and projections of how students are likely to score on

future tests, assuming average progress. To calculate an estimate of a teacher's effectiveness, EVAAS aggregates test score data of students from the teacher's classes.[263] SAS uses two value-added models in North Carolina: a multivariate response model and a univariate response model.

The multivariate model is used to calculate growth for English language arts and math teachers in grades 3-8 using the end-of-grade tests, which use the same scale from year to year. The model converts students' test scores to normal curve equivalents (NCEs) for analyses, compares current to previous year's NCEs to calculate student gains, and compares the gains to a base year benchmark to determine growth.[264] The model produces teacher "effects," or value-added, which are gains expressed as deviations, whether above or below, from the average gain for the district as a whole.[265]

The multivariate model, like other SAS EVAAS multivariate models,

- uses all available data for students and includes records of students even if some of the data is missing, estimating means for data sets with missing data based on correlations of current and previous test scores, and
- assumes that teacher effects accumulate over time.[266]

The univariate model is used for end-of-grade tests in grades 5 and 8 science, the end-of-course tests in high school, and the CTE Post-Assessments; it is proposed for the NC Final Exams. The univariate model uses a student's prior test scores to predict a student's performance on a certain assessment. To measure the progress of students on tests that are not administered in sequential years (all tests except math and reading), the model uses the scores of all the students who took the test plus their past testing history in other subjects. After students are tested, EVAAS determines the relationship between past testing histories of students in other subjects and their scores on the current year's test to determine the average progress for students with similar past testing histories. From this information, the model predicts where students should have scored, if they made average progress, and compares the predicted score to the actual score. This information is aggregated for all the students who took that test within a classroom to provide the value-added teacher effect. Conceptually, the value-added effect is the difference between the progress that groups of students made and the progress students similar to them made, on average, statewide.[267]

EVAAS does not control separately for student demographics or other characteristics, allowing students to serve as their own control over time.[268] The model addresses small class sizes by using a shrinkage method.[269] North Carolina requires a three-year average for teachers' growth model scores to help minimize year-to-year fluctuations.[270]

For each teacher with EVAAS scores, SAS reports a teacher estimate and a standard error. The teacher estimate is the combined growth of all the teacher's students who meet the minimum enrollment requirement (i.e., a student must be enrolled for 140 days for a year-long class or 70 days for a semester-long class at the time of test administration; students taking the test but who do not meet the enrollment requirements are not included in teacher value-added calculations).[271] The teacher estimate is expressed in NCE points; the standard error establishes a confidence band around the teacher estimate.[272]

SAS also provides a teacher index rating by dividing the teacher estimate by the standard error, reflecting the number of students in their classes and the statistical certainty around their estimate.[273] The key to the value-added levels (i.e., what the teacher index number means) is shown in Exhibit 4.

Exhibit 4: Key to Teacher Index

Effectiveness Level	Comparison to State Growth Standard/State Average
Exceeds Expected Growth	***Index is 2 or more***: teachers whose students are making substantially more progress than the state growth standard/state average
Meets Expected Growth	***Index is equal to or greater than -2 but less than 2***: teachers whose students are making the same amount of progress as the state growth standard/state average
Does Not Meet Expected Growth	***Index is less than -2***: teachers whose students are making substantially less progress than the state growth standard/state average.

Source: SAS Institute, http://si2013.ncdpi.wikispaces.net/file/view/Interpreting+the+Teacher+Value+Added+Report-MRM.pdf, North Carolina Department of Public Instruction (accessed July 25, 2014).

7. What issues and concerns does North Carolina face from using statistical growth model data in teacher evaluations?

No issues were identified beyond the difficulties of communicating complex measures to educators. As one research group noted, “Educators acknowledge that student growth is an important indicator of effectiveness, but their perspectives around the use of student growth data in evaluation reflects some confusion about the measures. Their misperceptions related to Standard 6 as a growth measure, coupled with uncertainties about the formulas used to calculate an effectiveness rating, raise concerns about the ability to effectively use student growth data to inform instruction.”[274]

OHIO

Background

Ohio received a $400 million Race to the Top grant in January 2011.[275] In its application, Ohio pledged to conduct annual teacher evaluations aligned to the state model, to train and credential the evaluators, and to use teacher evaluation data that includes student growth measures to design targeted supports and professional development and to make decisions on retention, dismissal, tenure, and compensation.

In November 2011, the Ohio State Board of Education approved a framework for a new teacher evaluation system that includes student growth and classroom observations. Local districts and charter schools that received Race to the Top funds were required to adopt the Ohio Teacher Evaluation System by July 2013 or to develop their own system aligned with the state board's framework. Districts have been allowed to phase in some portions of the new teacher evaluation system, but are expected to have fully implemented the new system in 2014-15.

Ohio was granted a No Child Left Behind waiver in May 2012; its waiver proposal addressed the continued implementation of its new teacher evaluation system. The state's one-year waiver extension was approved in August 2014.

1. Which teachers are required to have statistical growth model (value-added) data as part of their evaluations?

Teachers of math and reading in grades 4-8 whose students have taken the Ohio Achievement Assessments in the past are required to have growth model data included in their evaluations. The Ohio Achievement Assessments (OAA) are being replaced in 2014-15 for many grades and subjects with new assessments based on the Ohio New Learning Standards.

2. What weight is given to statistical growth model data in teacher evaluations?

In 2013-14, during the phase-in process of the new teacher evaluation system, growth model scores had to be weighted as at least 26 percent of teachers' evaluation scores with the expectation that districts would be required to weight them at the full 50 percent of evaluation components beginning in 2014-15.[276] In June 2014, the Ohio General Assembly revised several aspects of the teacher evaluation system based on feedback from educators. One change was to give local districts the option to reduce the weighting of growth model scores from 50 percent to 42.5 percent. The bill revising the teacher evaluation requirements originally sought to reduce the weighting to 35 percent; the final 42.5 percent was the compromise agreement.[277]

3. What measures are used in place of statistical growth model data for teachers of non-tested grades and subjects?

Districts can determine the student growth measures for teachers of non-tested grades and subjects. Districts may select from a state approved list of vendor-supplied assessments or may use local measures that meet state criteria. Teachers with a mix of classes, some with associated value-added data and some without, are rated on student growth using a mix of valued-added and district measures proportional to their class schedule.

The vendor assessments approved by the Ohio Department of Education may include nationally normed standardized assessments, industry certification examinations, or end-of-course examinations for which value-added data based on OAA is not available. If a district is already using, or chooses to use, state-approved vendor assessments, districts must weight these scores as at least 10 percent of teachers' student growth component. SAS Institute, Inc., Ohio's value-added provider, calculates student growth data for teachers from vendor assessments (including Terra Nova, ACT QualityCore EOC, NWEA MAP, and Renaissance Star) for the more than 100 Ohio districts that have opted to use them. (See more about SAS and value-added at Question 6.)

For most teachers (an estimated 70 percent), district-determined local measures will be used for the student growth component.[278] Student learning objectives (SLOs) are the primary local measures but school or district level value-added data may also be used, although the state recommends the value-added not be used for the entire student growth component.

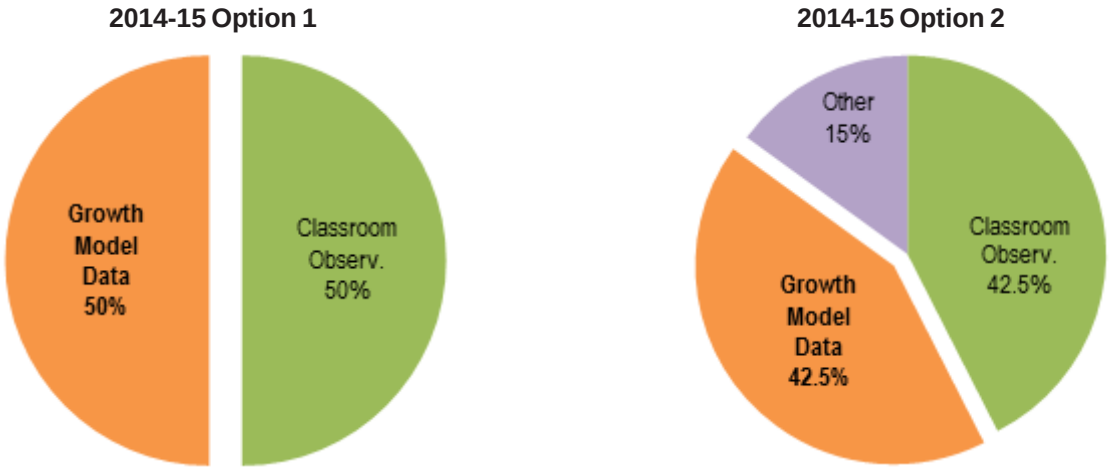
4. What other components are part of teacher evaluations in addition to student growth data?

All districts must include a measure of teachers' instructional practice (primarily classroom observations) as a component of equal weight to student growth data in their evaluations, but also have the option to include an additional "alternate component," which is the district's choice of student surveys, teacher self-evaluation, peer review evaluations, or student portfolios.

Under the Ohio Teacher Evaluation System in 2014-15, districts may weight the statistical growth model data at 50 percent and classroom observations for the other 50 percent of teachers' overall evaluation ratings as required under the original system (Exhibit 1, Option 1).[279] However, to incorporate legislative changes in 2014, districts also have the option to weight growth data at 42.5 percent, classroom observations at 42.5 percent, and attribute the remaining 15 percent to the alternate component (Exhibit 1, Option 2).[280] Beginning in 2015-16 districts must weight statistical growth model data and classroom observations equally, between 42.5 percent to 50 percent, with any remaining weight (1 percent to 15 percent) given to the alternate component chosen by the district.

Teacher instructional practice is determined through formal and informal classroom observations, as well as teacher conferences. Evaluators rate performance using the Teacher Performance Evaluation Rubric, based on the *Ohio Standards for the Teaching Profession*. Districts using the alternative component must select one of the department-approved instruments.

Exhibit 1: Ohio's Components of Teacher Evaluation



Note: Growth model data (value-added) weights are for teachers who instruct subjects with value-added data exclusively. The weighting is reduced for teachers who teach some value-added subjects on the proportion of those subjects in their teaching schedules.

5. How are teacher evaluation performance ratings applied to personnel decisions?

Overall teacher evaluation ratings are based on four performance levels that combine the instructional practice measure (classroom observation) and the alternative component ratings (both on a four-tier scale) plus the student growth rating, which is on a five-tier scale. (See more at Question 6 about the student growth rating.) Each rating is on a 600-point scale that is used to determine the overall (summative) evaluation score, as shown in Exhibit 2.

Exhibit 2: Combining Evaluation Components Option 2

Student Growth		Performance		Alternative Component		Final Summative Rating	
•Most Effective (5)	600	•Accomplished (4)	600	•Level 4	600	•Accomplished	500-600
•Above Average (4)	400	•Skilled (3)	400	•Level 3	400	•Skilled	300-499
•Average (3)	300	•Developing (2)	200	•Level 2	200	•Developing	100-299
•Approaching Average (2)	200	•Ineffective (1)	0	•Level 1	0	•Ineffective	0-99
•Least Effective (1)	0						

Note: A similar table exists for districts choosing a system with student growth and teacher performance each weighted 50 percent with no alternative component.
Source: Ohio Department of Education, Ohio Teacher Evaluation System Model, 2014, p.7, http://education.ohio.gov/ (accessed Oct. 29, 2014).

The state requires that each teacher have a professional growth or improvement plan, based on their student growth ratings. Teachers with top ratings (most effective) are allowed self-directed growth plans; those with average or above average growth ratings have collaborative growth plans; and those with low ratings (below average and approaching average) are required to have improvement plans directed by their evaluator. Teachers rated ineffective on their instructional practice component or their overall evaluation must also have improvement plans.

Legislation passed in 2014 allows districts to evaluate teachers less often than annually if their overall evaluation rankings are skilled or accomplished, and as long as their student growth rating is average or higher.[281]

Each local school board is to develop its own policies for using teacher evaluation results for retention, promotion, and dismissal. Seniority cannot be the basis for teacher retention decisions except among teachers with comparable evaluations. Districts can apply for a "safe harbor" provision to wait until 2015-16 to begin using evaluations that include state required value-added data to be used in dismissal, retention, tenure, or compensation decisions.

6. What type of statistical growth model is used for teacher evaluation?

Ohio has adopted SAS EVAAS as its statistical growth model. The state began pilot testing growth data for individual teachers in 2010-11; data was provided for all eligible teachers statewide in 2013. Ohio began calculating value-added data at the district and school levels in 2002 and has used the data as one of its school and district accountability measures since 2008.[282]

The EVASS model uses two approaches to calculate value-added scores for teachers depending on the type of tests used to assess student gains.

- A *multivariate* approach is used to calculate students' gains based on Ohio's reading and math end-of-grade, consecutive assessments given in grades 3-8. Statewide, all teachers whose overall evaluations must use value-added data for the student growth data component receive value-added data based on this approach.

- A *univariate* approach is used to calculate students' gains on assessments that may be administered in selected grades or selected subjects, such as the state science assessments in grades 5 and 8 and the district-selected vendor assessments (e.g., Terra Nova, ACT end-of-course assessments). This approach does not apply to teacher level value-added data required statewide for evaluation purposes.

The multivariate approach uses all available data for students (up to five prior years of test scores) and includes records of students even if some of the data for that student is missing.[283] Like other EVAAS models, it uses a correlation of current and previous scores to estimate means for data sets with missing data and assumes that teacher effects accumulate over time ("layering").[284] The Ohio model allows for re-estimation of teacher effects based on students' subsequent year achievement. The model converts test scale scores to normal curve equivalents (NCEs, which are similar to percentile rankings) to calculate student gains and compares them against students with similar achievement, using 2010 student results as the base year benchmark.[285] The Ohio multivariate approach estimates whether students maintained the same relative position (based on their NCE ranking) compared to the base year benchmark for the same subject and grade. Teachers' effects or impacts on student learning are their classrooms' deviation, whether up or down, from the average state gain for that grade, subject, and year.

The univariate approach is based on the difference between students' actual (observed) scores and their predicted scores, based on weighted composites of their previous test scores on other tests. Different composites with different weighting are used depending on the subject for which the estimated gain score is being computed. The model requires student records with at least three prior test scores in any subject or grade but may use up to five years of student data.[286] There is no "layering" of previous teachers' effects; prior school impact is considered part of the composite score. Teachers' effects or impacts are based on the growth of their students relative to the state average for classes in the same subject, grade, and year.

Like other EVAAS models, Ohio's does not include controls for classroom (peer) effects or school effects, and does not directly control for student demographics or other characteristics, allowing students to serve as their own control over time. Although not part of the formula for value-added, data on student characteristics are collected with student test data, which SAS uses in creating accountability reports for the public and web applications for educators. The data includes students' status as gifted, migrant, limited English proficient, economically disadvantaged, having disabilities, gender, and race.

SAS reports standard error and confidence intervals with teacher effect data, which is incorporated into an index score when included in teachers' overall evaluations. (Effect estimate divided by standard error = index score.) The model uses a shrinkage method to help ensure conservative estimates of teachers' effects, especially if they have small numbers of students. Teachers must have a minimum of six full-time-equivalent students in the same subject and grade for an individual value-added score to be calculated, and students absent from a class 45 or more days are not included in a teacher's value-added data. Ohio combines a teacher's value-added scores from all tested courses into a composite value-added score and combines individual year composite scores for up to a three-year average, when available, for value-added data used in evaluations.

Ohio uses a five-tier ranking to rate teachers' value-added data. (See Exhibit 2.) The rankings and associated growth, or index, scores are:[287]

Most Effective (5)	2.0 or above
Above Average (4)	At least 1.0 but lower than 2.0
Average (3)	Below 1.0 but not lower than -1.0
Approaching Average (2)	Below -1.0 but not lower than -2.0
Below Average (1)	Below -2.0

7. What are the issues and concerns facing Ohio from using statistical growth model data in teacher evaluations?

An analysis by state news media of 2011-12 growth model data found that schools and teachers with large percentages of poor students were more likely to have lower value-added scores than those serving more affluent students.[288] The Ohio Department of Education, in its own analysis of 2014

school level data, found that the statistical relationship between poverty in school and school value-added was very small; it had not analyzed the teacher level data.[289] Research has found that schools in low-income areas with high concentrations of students in poverty are more likely to have less experienced and less effective teachers due to several reasons, including higher turnover; it is not known if this was a factor in the news media findings.

Most of the impacts of teacher evaluation scores are determined locally by each district's policies. Recent legislative changes were implemented to provide a menu of options for districts to utilize.[290]

Endnotes

1 Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 10, http://www.nctq.org/ (accessed April 15, 2014).

2 U.S. Department of Education, "The No Child Left Behind Act of 2001," Jan. 7, 2002, p.1, http://www2.ed.gov/nclb/ (accessed Dec. 16, 2015).

3 Thomas B. Hoffer, E.C. Hedberg, Kevin L. Brown, et al., *Final Report on the Evaluation of the Growth Model Pilot Project*, U.S. Department of Education, 2011, p. 15, http://www2.ed.gov/ (accessed Dec. 15, 2014).

4 Wayne C. Riddle, *Adequate Yearly Progress (AYP): Growth Models under the No Child Left Behind Act*, Oct. 31, 2008, p.10, Congressional Research Service, http://congressionalresearch.com/ (accessed June 18, 2014).

5 Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, The New Teacher Project, 2009, pp. 6, 9, http://tntp.org/ (accessed Dec. 30, 2014).

6 Amy M. Hightower, Rachael C. Delgado, Sterling C. Lloyd, et al., *Improving Student Learning by Supporting Quality Teaching: Key Issues, Effective Strategies*, Editorial Projects in Education Research Center, 2011, p.5, http://www.edweek.org/ (accessed Dec. 5, 2014).

7 Pamela D. Tucker and James H. Stronge, "Chapter 1: The Power of an Effective Teacher and Why We Should Assess It," *Linking Teacher Evaluation and Student Learning*, 2005, http://www.ascd.org/ (accessed Dec. 4, 2014); Nancy Protheroe, *Improving Teaching and Learning with Data-based Decisions: Asking the Right Questions and Acting on the Answers*, Educational Research Service, 2001, http://www.rogersschools.net/ (accessed Dec. 4, 2014).

8 U.S. Department of Education, "President Obama, U.S. Secretary of Education Duncan Announce National Competition to Advance School Reform," http://www.ed.gov/ (accessed Dec. 16, 2014).

9 U.S. Department of Education, *Race to the Top Program – Executive Summary*, Nov. 2009, p. 3, http://www2.ed.gov/ (accessed Dec. 16, 2014).

10 Of the 485 points possible on the 19 Race to the Top grant criteria, the top point score was 65 points (13 percent) for "articulating state's education reform agenda and local educational agencies' participation in it." The second highest point score was 58 (12 percent) for teacher evaluation systems.

11 U.S. Department of Education, *Race to the Top Program – Executive Summary*, Nov. 2009, p. 9, http://www2.ed.gov/ (accessed Dec. 16, 2014).

12 Joseph C. Conaty, Interim Director, Race to the Top, U.S. Department of Education, letter addressed to Gov. Phil Bredesen, July 26, 2010, http://www2.ed.gov/ (accessed Dec. 30, 2014).

13 Offices of Research and Education Accountability, *Trends in Teacher Compensation: Focus on Alternative Salary Schedules*, Tennessee Comptroller of the Treasury, Jan. 2013, p. 9, http://www.comptroller.tn.gov/ (accessed Dec. 17, 2014).

14 U.S. Department of Education, "Teacher Incentive Fund FY 12 (Cohort 4) Awarded Grants," http://www2.ed.gov/ (accessed Dec. 4, 2014); Knox County Schools, "TAP expanded to 13 more Knox County Schools," http://highschools.knoxschools.org/ (accessed Aug. 21, 2012); Center for Education Compensation Reform, "TIF Grantees – Memphis City Schools," http://www.cecr.ed.gov/ (accessed Aug. 21, 2012) and TIF Grantees – State of Tennessee (Tennessee Department of Education) http://cecr.ed.gov/ (accessed Jan 2, 2014) .

15 U.S. Department of Education, "Principles for Improving Student Achievement and Increasing the Quality of Instruction," *ESEA Flexibility*, updated June 7, 2012, pp. 2-3, http://www.ed.gov/ (accessed Jan. 2, 2014).

16 *Education Week*, "NCLB Waivers: A State by State Breakdown," Nov. 13, 2014, http://www.edweek.org/ (accessed Dec. 30, 2014).

17 Alyson Klein, "Seven states put on NCLB waiver-renewal fast track," *Education Week*, Nov. 25, 2014, http://blogs.edweek.org/ (accessed Dec. 16, 2014).

18 Arne Duncan, Secretary, U.S. Department of Education, letter addressed to Chief State School Officers, June 18, 2013, http://www2.ed.gov/ (accessed Dec. 16, 2014).

19 Deborah S. Delisle, Assistant Secretary, U. S. Department of Education, letter addressed to Chief State School Officers, Aug. 21, 2014, http://www2.ed.gov/ (accessed Oct. 31, 2014).

20 Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 32, http://www.nctq.org/ (accessed April 15, 2014). Also see examples in this report's state profiles: District of Columbia (15 percent), North Carolina (40 percent), and Tennessee (44 percent).

21 Some states test reading comprehension, writing, and grammar separately or in various combinations. The term "English language arts" is used throughout this report to refer to all tests in these subject areas.

[22] Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 18, http://www.nctq.org/ (accessed April 15, 2014).

[23] Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, pp. 88-90, http://www.nctq.org/ (accessed April 15, 2014); Jim Hull, *Trends in Teacher Evaluation: How states are measuring teacher performance*, Center for Public Education, Oct. 2013, p. 12, www.centerforpubliceducation.org/ (accessed Sept. 16, 2014). At least one state, Tennessee, allows teachers to use their value-added score for 100 percent of their evaluation score if they so choose. (*Tennessee Public Acts* 2013, Ch. 105).

[24] Measures of Effective Teaching Project , *Ensuring Fair and Reliable Measures of Effective Teaching*, Bill and Melinda Gates Foundation, Jan. 2013, p. 5, http://www.metproject.org/ (accessed Dec. 17, 2014).

[25] Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 18, http://www.nctq.org/ (accessed April 15, 2014).

[26] Reform Support Network, *The View from the States: A Brief on Non-tested Grades and Subjects*, U.S. Department of Education, [no date], p.6, http://www2.ed.gov/ (accessed Dec. 17, 2014); Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. vi, http://www.nctq.org/ (accessed April 15, 2014).

[27] Policies of the Tennessee State Board of Education, 5.201, Teacher and Principal Evaluation Policy, p. 2, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014).

[28] Measures of Effective Teaching Project, *Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project*, Bill and Melinda Gates Foundation, 2010, p.9, http://www.metproject.org/ (accessed Dec. 18, 2014).

[29] Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 30, http://www.nctq.org/ (accessed April 15, 2014).

[30] Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. 13, http://www.nctq.org/ (accessed April 15, 2014).

[31] Kathryn M. Doherty and Sandi Jacobs, *State of the States 2013 – Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice*, National Council on Teacher Quality, Oct. 2013, p. ii, http://www.nctq.org/ (accessed April 15, 2014).

[32] Hua Wei, Tracey Hembry, Daniel L. Murphy, Yuanyuan McBride, *Value-added Models in the Evaluation of Teacher Effectiveness: A comparison of Models and Outcomes*, Pearson Research Report, May 2012, p.3; Dan Goldhaber and Roddy Theobald, *Do Different Value-added Models Tell Us the Same Things?*, Carnegie Knowledge network, updated April 15, 2013, pp. 1, 6, 7, www.carnegieknowledgenetwork.org/ (accessed Oct. 11, 2014); American Statistical Association, *ASA Statement on Using Value-added Models for Educational Assessment*, April 8, 2014, p.2, http://www.amstat.org/ (accessed April 9, 2014).

[33] Dan Goldhaber, Joe Walch, and Brian Gabele, *Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments*, Center for Education Data and Research, Aug. 10, 2012, pp. 6, 27, http://www.cedr.us/ (accessed April 17, 2014).

[34] TVAAS reports for schools and districts also show three-year averages, as well as single-year growth.

[35] Other issues of standardized testing raised in the context of teacher evaluation – such as tests not measuring higher-order skills that teachers are teaching (logical thinking, analyzing abstract problems), teachers narrowing the curriculum by increasing emphasis on tested concepts and subjects, students' test scores rising from class focus on test preparation and test-taking skills without students actually learning more (score inflation), and the distortion of test results as more high stakes decisions are linked to them – are not directly related to the use of growth models and are beyond the scope of this report.

[36] SAS Institute, Inc., *Misconceptions about Value-added Reporting in Tennessee*, 2012, p.13, http://www.tn.gov/ (accessed Nov. 12, 2014).

[37] John P. Papay, "Different Tests, Different Answers: The Stability of Teacher value-added Estimates Across Outcome Measures," *American Educational Research Journal*, April 19, 2010, http://aer.sagepub.com/ (accessed Oct. 11, 2014); Tim R. Sass, *The Stability of Value-added Measures of Teacher Quality and Implications for Teacher Compensation Policy*, National Center for Analysis of Longitudinal Data in Education Research, Nov. 2008, p.5, http://www.urban.org/ (accessed Nov. 12, 2014).

[38] Dan Goldhaber, Joe Walch, and Brian Gabele, *Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments*, Center for Education Data and Research, Aug. 10, 2012, p. 27, http://www.cedr.us/ (accessed April 17, 2014); Susanna Loeb and Christopher A. Candelaria, *How Stable are Value-added Estimates Across Years, Subjects, and Student Groups?*, Carnegie Knowledge Network, revised July 2013, pp. 3, 8, www.carnegieknowledgenetwork.org/ (accessed Oct. 10, 2014); Tim R. Sass, *The Stability of Value-added Measures of Teacher Quality and Implications for Teacher Compensation Policy*, National Center for Analysis of Longitudinal Data in Education Research, Nov. 2008, p.2, http://www.urban.org/ (accessed Nov. 12, 2014); Office of Education Accountability, *The Measure of Education: A Review of the Tennessee Value-added Assessment System,* Tennessee Comptroller of the Treasury, April 1995, p. 11, http://www.comptroller.tn.gov/ (accessed Dec. 15, 2014).

[39] Alyson Leah Lavigne and Thomas L. Good, *Evaluating Teachers and Students: Moving Beyond the Failure of School Reform*, 2013, New York: Routledge, p. 133; SAS Institute, Inc., *Misconceptions about Value-added Reporting in Tennessee*, 2012, p.17, http://www.tn.gov/ (accessed Nov. 12, 2014).

[40] SAS Institute, Inc., *Misconceptions about Value-added Reporting in Tennessee*, 2012, p.17, http://www.tn.gov/education/ (accessed Nov. 12, 2014).

[41] Douglas N. Harris, William K. Ingle, and Stacey A. Rutledge, "How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-added Measures," *American Education Research Journal*, Vol. 51, No. 1, Feb. 2014, p. 94, http://aer.sagepub.com/ (accessed July 21, 2014).

[42] SAS Institute, Inc., *Misconceptions about Value-added Reporting in Tennessee*, 2012, p.17, http://www.tn.gov/education/ (accessed Nov. 12, 2014).

[43] Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014. Data reflects Tennessee Department of Education analysis of TVAAS data from 2011-12 and 2012-13.

[44] Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014. Data reflects Tennessee Department of Education analysis of TVAAS data from 2011-12 and 2012-13.

[45] Battelle for Kids, *Selecting Growth Measures: A Guide for Education Leaders*, 2011, p.10, http://www.edgrowthmeasures.org/ (accessed Dec 16, 2014).

[46] *Tennessee Code Annotated* 49-1-606(a).

[47] Douglas N. Harris, William K. Ingle, and Stacey A. Rutledge, "How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-added Measures," *American Education Research Journal*, Vol. 51, No. 1, Feb. 2014, p. 73, http://aer.sagepub.com/ (accessed July 21, 2014).

[48] Douglas N. Harris, William K. Ingle, and Stacey A. Rutledge, "How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-added Measures," *American Education Research Journal*, Vol. 51, No. 1, Feb. 2014, pp. 89-92, http://aer.sagepub.com/ (accessed July 21, 2014).

[49] Morgan S. Polikoff and Andrew C. Porter, "Instructional Alignment as a Measure of Teaching Quality," Educational Evaluation and Policy Analysis, May 12, 2014, p. 11-14, http://epa.sagepub.com/ (accessed June 20, 2014).

[50] Michael Hansen, Mariann Lemke, Nicholas Sorensen, *Combining Multiple Performance Measures: Do Common Approaches Undermine Districts' Personnel Evaluation Systems?*, American Institutes for Research, Aug. 2013, p. 2, http://www.air.org/ (accessed Aug. 8, 2014). The Gates' Measures of Effective Teaching study produced several reports. See, for example, *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*, 2010; this and other project reports can be found at http://www.metproject.org/.

[51] Measures of Effective Teaching Project, Gathering Feedback for Teaching, Bill and Melinda Gates Foundation, Jan. 2012, p. 9, http://www.metproject.org/ (accessed Dec. 18, 2014).

[52] Holly Yettick, "Studies Highlight Complexities of Using Value-added Measures," *Education Week*, May 13, 2014, http://www.edweek.org/ (accessed May 14, 2014).

[53] Complaint filed March 19, 2014, U.S. District Court, Eastern Tennessee, Taylor v. Haslam, Huffman, and Knox County Board of Education, Case 3:14-cv-00113-PLR-CCS; Amended complaint filed March 5, 2014, U.S. District Court, Eastern Tennessee, Trout v. Knox County Board of Education, Case 3:14-cv-00049-HSM-CCS; and various news articles: Hailey Heinz, "Suit filed to stop teacher evaluation," *Albuquerque Journal*, Sept. 12, 2013, http://www.abqjournal.com/ (accessed Aug. 12, 2014); Valerie Strauss, "High-achieving teacher sues state over evaluation labeling her 'ineffective'," *Washington Post*, Oct. 31, 2014, http://www.washingtonpost.com/ (accessed Dec. 19, 2014); Valerie Strauss, "Houston teachers sue over controversial teacher evaluation method," *Washington Post*, April 30, 2014, http://www.washingtonpost.com/ (accessed May 6, 2014).

[54] Alyson Klein and Andrew Ujifusa, "Waiver-renewal Twist Centers on Teacher Evaluations," Education Week, May 20, 2014, http://www.edweek.org/ (accessed

May 21, 2014). The other states cited as at the forefront include New Jersey, New Mexico, and New York.

Tennessee Profile

[55] 106th Tennessee General Assembly, Public Acts 2010, First Extraordinary Session, Chapter 2, First to the Top Act of 2010, http://state.tn.us/sos/ (accessed Dec. 19, 2014).

[56] Joseph C. Conaty, Interim Director, Race to the Top, U.S. Department of Education, letter, addressed to Gov. Phil Bredesen, July 26, 2010, http://www2.ed.gov/ (accessed Dec. 30, 2014).

[57] Deborah S. Delisle, Assistant Secretary, U.S. Department of Education, letter, addressed to Commissioner Kevin Huffman, Tennessee Department of Education, Sept. 5, 2014, http://www2.ed.gov/ (accessed Dec. 8, 2014).

[58] Student scores from Advanced Placement and International Baccalaureate course exams in comparable high school subject areas are not included in TVAAS calculations for individual teachers or school-wide value-added scores.

[59] 108th Tennessee General Assembly, Public Acts 2013, Chapter 105, An Act to amend Tennessee Code Annotated, Section 49-1-302(d) and Section 49-1-606(a), relative to teacher and principal evaluations, http://www.tn.gov/sos/ (accessed Nov. 7, 2014); Luke Kohlmoos, Director of Evaluation, Tennessee Department of Education, e-mail, Nov. 26, 2014.

[60] Michael McWeeney, TEAM Program Analyst, Tennessee Department of Education, e-mail, Feb. 5, 2015, and telephone interview, Feb. 9, 2015. Percentages are based on the pool of teachers for whom complete evaluation data is on file.

[61] *Tennessee Code Annotated* 49-1-302(d)(2)(B)(ii).

[62] Policies of the Tennessee State Board of Education, 5.201, Teacher and Principal Evaluation Policy, p.3, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014); 108th Tennessee General Assembly, Public Acts 2013, Chapter 105, An Act to amend Tennessee Code Annotated, Section 49-1-302(d) and Section 49-1-606(a), relative to teacher and principal evaluations, http://www.tn.gov/sos/ (accessed Nov. 7, 2014).

[63] *Tennessee Code Annotated* 49-1-302(d)(2)(B)(vi).

[64] Policies of the Tennessee State Board of Education, 5.201, Teacher and Principal Evaluation Policy, p. 2, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014).

[65] Policies of the State Board of Education, 5.201, Teacher and Principal Evaluation Policy, p. 2, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014).

[66] *Tennessee Code Annotated* 49-1-302(d)(2)(B)(iii).

[67] The law states that if the teacher and person responsible for conducting the evaluation do not agree on the measures to use for the student achievement component, then the teacher makes the final decision.

[68] *Tennessee Code Annotated* 49-1-302(d)(2)(B)(i-iii) and (vi).

[69] *Tennessee Code Annotated* 49-1-302(d)(2)(C).

[70] TAP is the name (not an acronym) for The System for Teacher and Student Advancement, a comprehensive approach to school reform based on multiple career paths for teachers, ongoing professional development, classroom evaluation, and performance-based compensation. Tennessee has adopted only the classroom evaluation rubric of this system.

[71] Policies of the Tennessee State Board of Education, 5.201, Teacher and Principal Evaluation Policy, pp. 6-8, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014).

[72] Ibid., p. 1.

[73] *Tennessee Code Annotated* 49-1-302 (d)(2)(A).

[74] Tennessee Department of Education, *Redesigning the Pathway to Teaching*, Spring 2013, pp. 14-17, presented at the June 21, 2013 State Board Workshop; Tennessee State Board of Education, Minutes, Conference Call Meeting, Item II E, Aug. 16, 2013, pp. 4-6; http://www.tn.gov/ (accessed Dec. 10, 2014).

[75] Tennessee Department of Education, "Changes to Licensure Policy in 2015," http://www.state.tn.us/ (accessed Dec. 10, 2014); 108th General Assembly, Public Acts, 2014, Ch. 746, An Act to amend Tennessee Code Annotated, Title 49, Chapter 1; Title 49, Chapter 2 and Title 49, Chapter 5, Part 1, relative to licensure, http://www.tn.gov/sos/ (accessed Dec. 10, 2014).

[76] *Tennessee Code Annotated* 49-5-503(4).

[77] *Tennessee Code Annotated* 49-5-501(11)(B)(i).

[78] *Tennessee Code Annotated* 49-5-504(e) and (f).

[79] *Tennessee Code Annotated* 49-3-306(h).

[80] Tennessee Department of Education, "Differentiated Pay," http://www.tn.gov/ (accessed Dec. 9, 2014).

[81] *Tennessee Code Annotated* 49-5-511(a)(2).

[82] Level 1 is defined as significantly below expectations and Level 2 is below expectations. 107th General Assembly, Public Acts, 2011, Chapter 70, An Act to amend Tennessee Code Annotated, Title 49, Chapter 5, relative to the employment of teachers, http://www.tn.gov/sos/ (accessed Nov. 7, 2014).

[83] 108th General Assembly, Public Acts, 2013, Chapter 369, An Act to amend Tennessee Code Annotated, Section 49-5-511, relative to dismissal or suspension of teachers, http://www.tn.gov/sos/ (accessed Sept. 23, 2014).

[84] *Tennessee Code Annotated* 49-1-302 (d)(2 (A).

[85] Policies of the Tennessee State Board of Education, 5.201, Teacher and Principal Evaluation Policy, p. 10, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014).

[86] Ibid., p. 9.

[87] SAS Institute, Inc. markets the Education Value-Added Assessment System (EVAAS), which builds on the TVAAS methodology developed by Dr. Sanders directly from TVAAS.

[88] 106th Tennessee General Assembly, Public Acts 2010, First Extraordinary Session, Chapter 2, First to the Top Act of 2010, sec. 10 (d)(2)(A)(i), http://state.tn.us/sos/ (accessed Dec. 19, 2014). The Education Improvement Act (1992) required the State Board of Education to set guidelines for local districts' teacher evaluation policies that would include classroom observations, review of prior evaluations and personal conferences plus "other appropriate criteria including the Sanders model" data. State Board policy adopted in 1996 clarified that, pending research findings, TVAAS data was to be used only for diagnostic purposes. The Framework for Evaluation and Professional Growth (FEPG) was implemented statewide in 2000, and indicated that individual teacher TVAAS data, if available, was to be part of the evaluation documentation, and could be used to develop a teacher's required professional development plan.

[89] Office of Education Accountability, *The Measure of Education: A Review of the Tennessee Value Added Assessment System*, Tennessee Comptroller of the Treasury, April 1995, p.4, http://www.comptroller.tn.gov/ (accessed Dec. 12, 2014); Tennessee Department of Education, "Tennessee Value-added Assessment System (TVAAS): A Conceptual Understanding" (video at "How TVAAS Works"), http://www.tn.gov/ (accessed Sept. 19, 2014).

[90] Tennessee Department of Education, "TVAAS FAQs," http://team-tn.cloudapp.net/ (accessed Dec. 12, 2014). Note that eligible subjects are those taught in the current year that have also been taught in the previous two years, regardless of grade level.

[91] *Tennessee Code Annotated* 49-1-603(b).

[92] 108th Tennessee General Assembly, Public Acts 2013, Chapter 105, An Act to amend Tennessee Code Annotated, Section 49-1-302(d) and Section 49-1-606(a), relative to teacher and principal evaluations, http://www.tn.gov/sos/ (accessed Nov. 7, 2014).

[93] *Tennessee Code Annotated* 49-1-606(b).

[94] *Tennessee Code Annotated* 49-1-603(b).

[95] William L. Sanders and Sandra P. Horn, "Research Findings from the Tennessee Value-added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education*, 1998, p.2, https://www.sas.com/ (accessed Dec 29, 2014).

[96] William L. Sanders, Arnold M. Saxton, and Sandra P. Horn, Chapter 13, "The Tennessee Value-added Assessment System: A Quantitative Outcomes-based Approach to Educational Assessment," in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, ed. Jason Millman, California: Corwin Press, Inc., 1997, pp. 137-144.

[97] Tennessee Department of Education, "TVAAS Tennessee Value-added Assessment System, Fact Sheet, Assessment Literacy Training," [no date], http://www.scsk12.org/ (accessed Sept. 8, 2014).

[98] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, pp. 2-3, March 25, 2010, http://www.sas.com/ (accessed Dec. 18, 2014).

[99] Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014.

[100] Jennifer L. Green, *Estimating Teacher Effects Using Value-added Models*, University of Nebraska at Lincoln, Department of Statistics: Dissertations and Theses, Aug. 1, 2010, pp. 24-28, http://digitalcommons.unl.edu/ (accessed Dec. 12, 2014); *Recommendations of the Florida Student Growth Implementation Committee: Background and Summary*, 2011, pp. 6, 8, http://www.fldoe.org/ (select Value-added Model While Paper) (accessed July 11, 2014).

[101] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, pp. 12-13, March 25, 2010, http://www.sas.com/ (accessed Dec. 18, 2014). Value-added for teachers in those districts that use K-2 testing is also calculated using the univariate method.

[102] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, p. 13, March 25, 2010, http://www.sas.com/ (accessed Dec. 18, 2014).

[103] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, p. 7, March 25, 2010, http://www.sas.com/ (accessed Dec. 18, 2014).

[104] *Tennessee Code Annotated* 49-1-603(b) and 49-1-604; Office of Education Accountability, *Education Improvement Act: A Progress Report*; Tennessee Comptroller of the Treasury, 2004, p. C-64, http://www.comptroller.tn.gov/ (accessed July 25, 2014).

[105] Tennessee Senate Education Committee, March 24, 2014, discussion of Senate Bill 2342, video clip 1:09:25 through 1:15:00.

[106] 108th Tennessee General Assembly, Public Acts, 2014, Chapter 796, An Act to amend Tennessee Code Annotated, Title 49, relative to the evaluations of teachers and principals, http://www.tn.gov/sos/ (accessed Jan 7, 2014).

[107] Tennessee Senate Education Committee, March 24, 2014, discussion of Senate Bill 2342, video clip 1:09:25 through 1:15:00.

[108] Policies of the Tennessee State Board of Education, 5.201, Teacher and Principal Evaluation Policy, p. 1, revised July 25, 2014, http://www.tn.gov/sbe/ (accessed Oct 13, 2014).

[109] Sara Heyburn, Assistant Commissioner, Teachers and Leaders, Tennessee Department of Education, "Request for Information," e-mail, Sept. 25, 2014.

[110] Tennessee Comptroller's Offices of Research and Education Accountability (OREA) analysis of teacher evaluation data provided by the Department of Education. See OREA's separate report on its analysis, *An Analysis of Teacher Evaluation Data and Teacher Characteristics,* http://www.comptroller.tn.gov/. The data comprises all educators who were evaluated as teachers which can include school counselors, social workers, or instructional coaches who teach part time, in addition to regular, full time classroom teachers.

[111] Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014. The school with the largest discrepancy had 78.6 percent of teachers with a three or more point gap between their classroom observation score and TVAAS score.

[112] Tony Pratt, Deputy Assistant Commissioner, Data and Research, Tennessee Department of Education, interview, Sept. 29, 2014; Luke Kohlmoos, Director of Evaluation, Tennessee Department of Education, e-mail, Nov. 26, 2014.

[113] Complaint filed March 19, 2014, U.S. District Court, Eastern Tennessee, Taylor v. Haslam, Huffman, and Knox County Board of Education, Case 3:14-cv-00113-PLR-CCS; Amended complaint filed March 5, 2014, U.S. District Court, Eastern Tennessee, Trout v. Knox County Board of Education, Case 3:14-cv-00049-HSM-CCS; Tennessee Education Association, "TEA Files Second TVAAS Lawsuit, Gov. and Commissioner Named as Defendants," [no date], http://www.teateachers.org/ (accessed Dec. 11, 2014).

Colorado Profile

[114] Colorado General Assembly, Senate Bill 10-191, http://tornado.state.co.us/ (accessed Sept. 30, 2014).

[115] Department of Education, Colorado State Board of Education, Rules for Administration of a Statewide System to Evaluate the Effectiveness of Licensed Personnel Employed by School Districts and Boards of Cooperative Services, 1 CCR 301-87, 5.01 (E)(7), p. 17, http://www.sos.state.co.us/ (accessed June 12, 2014). Using Race to the Top funds, the CDE created the Colorado Content Collaboratives, groups of Colorado educators who review and create high-quality assessments aligned to the state's academic standards and which may be used for educator evaluation purposes. Colorado Department of Education, Colorado Content Collaboratives, http://www.cde.state.co.us/ContentCollaboratives (accessed June 13, 2014).

[116] The General Assembly of the State of Colorado, Senate Bill 14-165, http://www.leg.state.co.us/ (accessed June 13, 2014); Colorado Department of Education, Partnerships for Assessment of Readiness for College and Careers (PARCC), http://www.cde.state.co.us/ (accessed June 13, 2014); Michael Johnston and Kerrie Dallman, "Giving Colorado's Teacher Evaluation Bill Time to Succeed," *The Denver Post*, March 16, 2014, http://www.denverpost.com/ (accessed Nov. 18, 2014).

[117] Colorado Department of Education, Colorado's Race to the Top Phase 3 Grant Program, http://www.cde.state.co.us/rttt (accessed June 13, 2014).

[118] Colorado Department of Education, ESEA Flexibility Request submitted to U.S. Department of Education, Jan. 20, 2011, http://www2.ed.gov/ (accessed Dec. 11, 2014).

[119] Deborah S. Delisle, Assistant Secretary, U.S. Department of Education, letter to Robert Hammond, Commissioner, Colorado Department of Education, July 3, 2014, http://www2.ed.gov/ (accessed Sept. 30, 2014).

[120] Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail and attachment, Sept. 29, 2014. Department of Education, Colorado State Board of Education, Rule 5.01 (E)(7) requires that the CGM results be used for the academic growth portion of a teacher's evaluation for subjects with annual statewide summative assessment results available in two consecutive grades. See http://www.sos.state.co.us/ (accessed June 12, 2014).

[121] Colorado Department of Education, Partnership for Assessment of Readiness for College and Careers (PARCC), http://www.cde.state.co.us/ (accessed Sept. 30, 2014).

[122] Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail, July 2, 2014.

[123] Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail, July 2, 2014; Colorado Department of Education, Frequently Asked Questions, "Senate Bill 10-191: 2014-15" and "Measuring the 50 Percent Measures of Student Learning (Growth) in Evaluations," http://www.cde.state.co.us/ (accessed Nov. 17, 2014).

[124] The General Assembly of the State of Colorado, Senate Bill 14-165, http://www.leg.state.co.us/ (accessed June 13, 2014).

[125] Colorado Department of Education, Measures of Student Learning Guidance for Districts: Version 2.0, Aug. 1, 2013, pp. 1-2 and 10, http://www.cde.state.co.us/ (accessed June 10, 2014); Colorado Department of Education, Determining Final Effectiveness Ratings Using the Colorado State Model Evaluation System for Teachers, (no date), p. 3, http://www.cde.state.co.us/ (accessed June 9, 2014).

[126] Colorado Department of Education, Colorado Content Collaboratives, http://www.cde.state.co.us/ (accessed June 13, 2014).

[127] Colorado Department of Education, Determining Final Effectiveness Ratings for Combining Professional Practices and Measures of Student Learning, (no date), pp. 1-2, 7, http://www.cde.state.co.us/ (accessed June 9, 2014); Colorado Department of Education, Rubric for Evaluating Colorado Teachers, 2014-15, http://www.cde.state.co.us/ (accessed Dec. 11, 2014).

[128] Colorado Department of Education, Determining Final Effectiveness Ratings for Combining Professional Practices and Measures of Student Learning, (no date), p. 1, http://www.cde.state.co.us/sites/ (accessed June 9, 2014).

[129] Colorado Department of Education, CDE Fact Sheet, Connecting Teacher Status and S.B. 10-191, Understanding the Non-probationary Implications of S.B. 10-191, Nov. 2013, http://www.cde.state.co.us/ (accessed Sept. 30, 2014); The Colorado State Board of Education, Rules for Administration of a Statewide System to Evaluate the Effectiveness of Licensed Personnel Employed by School Districts and Boards of Cooperative Services, 1 CCR 301-87, 3.03(D)(2)(a)(ii), pp. 11-12, http://www.sos.state.co.us/ (accessed Sept. 23, 2014); Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail attachment, July 2, 2014.

[130] Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail and attachment, Sept. 29, 2014.

[131] Colorado Department of Education, *2014-15 User's Guide, Colorado State Model Educator Evaluation System*, pp. 17-18, http://www.cde.state.co.us/ (accessed Sept. 30, 2014); Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail and attachment, Sept. 29, 2014.

[132] Colorado state law requires that the Colorado Growth Model "shall be a student longitudinal academic growth model that is available in the public domain." Dr. Damian Betebenner developed the CGM for the Colorado Department of Education. Colorado Revised Statutes 22-11-202. The model is also used by several other states, including Hawaii, Massachusetts, New Jersey, New York, and Rhode Island. Elias Walsh and Eric Isenberg, *How Does a Value-Added Model Compare to the Colorado Growth Model?*, Mathematica Policy Research, Oct. 2013, pp. 6 and 20, http://www.mathematica-mpr.com/ (accessed June 11, 2014); Damian Betebenner, An R Package for the Calculation and Visualization of Student Growth Percentiles & Percentile Growth Trajectories, March 5, 2014, http://cran.r-project.org/ (accessed June 11, 2014).

[133] Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail attachment, July 2, 2014.

[134] Elena K. Diaz-Bilello and Derek C. Briggs, *Using Student Growth Percentiles for Educator Evaluations at the Teacher Level: Key Issues and Technical Considerations for School Districts in Colorado*, Center for Assessment, July 2014, p. 11, http://www.cde.state.co.us/ (accessed Sept. 23, 2014); Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail, Sept. 29, 2014.

[135] Colorado Department of Education, 2009 Sample Growth and Achievement Report, http://www2.cde.state.co.us/ (accessed June 13, 2014).

[136] R. Katherine D. Castellano, Andrew D. Ho, A Practitioner's Guide to Growth Models, Council of Chief State School Officers, Feb. 2013, pp. 89-90, http://scholar.harvard.edu/ (accessed June 16, 2014).

[137] Colorado Department of Education, The Colorado Growth Model, Frequently Asked Questions, Updated July 2009, http://assessment.aurorak12.org/ (accessed Nov. 19, 2014); Elias Walsh and Eric Isenberg, *How Does a Value-Added Model Compare to the Colorado Growth Model?*, Mathematica Policy Research, Oct. 2013, p. 1, http://www.mathematica-mpr.com/ (accessed June 11, 2014).

[138] Cassandra M. Guarino, Mark D. Reckase, Brian W. Stacy, Jeffrey M. Wooldridge, *A Comparison of Growth Percentile and Value-Added Models of Teacher Performance*, Working Paper #39,

The Education Policy Center at Michigan State University, Feb. 6, 2014, pp. 6-8, http://education.msu.edu/ (accessed June 11, 2014).

[139] Elias Walsh and Eric Isenberg, *How Does a Value-Added Model Compare to the Colorado Growth Model?*, Mathematica Policy Research, Oct. 2013, p. 7, http://www.mathematica-mpr.com/ (accessed June 11, 2014).

[140] The Colorado State Board of Education, Rules for Administration of a Statewide System to Evaluate the Effectiveness of Licensed Personnel Employed by School Districts and Boards of Cooperative Services, 1 CCR 501-87, 5.01(E)(7)(d), p. 18, http://www.sos.state.co.us/ (accessed Sept. 23, 2014). However, the business rules for using Median Growth Percentiles in teacher evaluation are determined at the local level. Also, the two years do not necessarily need to be from the same assessment; if the alignment between two assessments is high enough, it may be possible and valid to measure growth between the two different assessments. Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail and attachment, Sept. 29, 2014.

[141] Elias Walsh and Eric Isenberg, *How Does a Value-Added Model Compare to the Colorado Growth Model?*, Mathematica Policy Research, Oct. 2013, p. 7, http://www.mathematica-mpr.com/ (accessed June 11, 2014).

[142] Elena K. Diaz-Bilello and Derek C. Briggs, *Using Student Growth Percentiles for Educator Evaluations at the Teacher Level: Key Issues and Technical Considerations for School Districts in Colorado*, Center for Assessment, July 2014, http://www.cde.state.co.us/ (accessed Sept. 23, 2014); Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail and attachment, Sept. 29, 2014.

[143] Toby King, Director of Evaluation and Support, Colorado Department of Education, e-mail and attachment, Sept. 29, 2014, and e-mail, Sept. 30, 2014.

[144] Colorado Revised Statutes 22-11-202.

[145] Technical Advisory Panel for Longitudinal Growth, Colorado Growth Model and Educator Effectiveness: Implementation Recommendations, Feb. 13, 2014, http://sites.cde.state.co.us/ (accessed Sept. 30, 2014).

[146] The General Assembly of the State of Colorado, Senate Bill 14-165, http://www.leg.state.co.us/ (accessed June 13, 2014).

District of Columbia Public Schools Profile

[147] Office of State Superintendent of Education, *Teacher Equity Plan*, 2013, pp. 1-2, http://osse.dc.gov/ (accessed June 25, 2014).

[148] District of Columbia Public Schools, "Chancellor Henderson Announces DCPS Will Not Use Student Test Scores in Teacher Evaluations Next School Year," June 19, 2014, http://dcps.dc.gov/ (accessed June 23, 2014).

[149] District of Columbia Public Schools, *Lessons Learned from Four Years of Implementation in DC Public Schools: An Overview*, May 28, 2014, p. 32, http://relnw.educationnorthwest.org/ (accessed Oct 8, 2014).

[150] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 1, General Education Teachers with Individual Value-Added Student Achievement Data, 2013-14, p.6, http://dcps.dc.gov/ (accessed Dec. 15, 2014).

[151] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 1, General Education Teachers with Individual Value-Added Student Achievement Data, 2012-13 p.3, http://dcps.dc.gov/ (accessed Dec. 14, 2014).

[152] PARCC (Partnership for Assessment of Readiness for College and Careers) will replace DC CAS in 2014-15. DCPS students in grades 3-8 and high school students in selected subjects will take the PARCC Assessment for English language arts and math.

[153] District of Columbia Public Schools, "Chancellor Henderson Announces DCPS Will Not Use Student Test Scores in Teacher Evaluations Next School Year," June 19, 2014, http://dcps.dc.gov/ (accessed June 23, 2014).

[154] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 2, Grades 1–12 General Education Teachers Without Individual Value-Added Student Achievement Data, 2013-14, p.6, http://dcps.dc.gov/ (accessed June 19, 2014).

[155] Different versions of the evaluation system have been developed for selected groups of teachers, such as for teachers of English Language Learners and students with special education needs.

[156] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 1, General Education Teachers with Individual Value-Added Student Achievement Data, 2013-14, pp.6,52,62, http://dcps.dc.gov/ (accessed Dec. 15, 2014); District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 2, Grades 1–12 General Education Teachers Without Individual Value-Added Student Achievement Data, 2013-14, pp.6, 48, 59, http://dcps.dc.gov/ (accessed June 19, 2014).

[157] TAS is based on appropriate assessments for the grade and subject area. Goals are unique and developed based on baseline data, Principals and teachers determine the most appropriate assessment and goals at the beginning of the school year. Teachers collect data over the course of the year and meet with the principal again at the end of the year to discuss and assign a final TAS score.

[158] District of Columbia Public Schools, "Overview of Impact," [no date], http://dcps.dc.gov/ (accessed June 19, 2014).

[159] Chad Aldeman and Carolyn Chuong, T*eacher Evaluations in an Era of Rapid Change: from "Unsatisfactory" to "Needs Improvement,"* Bellwether Education Partners, Aug. 2014, http://bellwethereducation.org/ (accessed Sept. 16, 2014); Sally Holland, "D.C. school system fires 241 teachers," *CNN U.S.*, July 23, 2010, http://www.cnn.com/ (accessed Oct. 6, 2014); Emma Brown, "98 D.C. teachers fired for poor performance," *The Washington Post*, Aug. 1, 2012, http://www.washingtonpost.com/ (accessed Dec. 14, 2014).

[160] District of Columbia Public Schools, *Lessons Learned from Four Years of Implementation in DC Public Schools: An Overview*, May 28, 2014, p. 10, http://relnw.educationnorthwest.org/ (accessed Oct 8, 2014).

[161] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 1, General Education Teachers with Individual Value-Added Student Achievement Data, 2013-14, pp.8-10, http://dcps.dc.gov/ (accessed Dec. 15, 2014).

[162] Eric Isenberg and Elias Walsh, *Measuring Teacher Value Added in DC, 2012–2013 School Year: Final Report*, Mathematica Policy Research, Sept. 11, 2013, pp. 5, 22, http://dcps.dc.gov/ (accessed June 18, 2014).

[163] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 1, General Education Teachers with Individual Value-Added Student Achievement Data, 2013-14, p.9, http://dcps.dc.gov/ (accessed Dec. 15, 2014).

[164] Alden Wells, Manager, IMPACT, and Luke Hostetter, Coordinator, IMPACT, District of Columbia Public Schools, telephone interview, Oct. 14, 2014 and e-mail attachment, Feb. 11, 2015.

[165] Alden Wells, Manager, IMPACT, and Luke Hostetter, Coordinator, IMPACT, District of Columbia Public Schools, telephone interview, Oct. 14, 2014. DCPS staff indicated that the district is considering the future possibility of using a value-added score based on an average of multiple years.

[166] District of Columbia Public Schools, *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*, Guidebook for Group 1, General Education Teachers with Individual Value-Added Student Achievement Data, 2013-14, p.11, http://dcps.dc.gov/ (accessed Dec. 15, 2014).

[167] Alden Wells, Manager, IMPACT, and Luke Hostetter, Coordinator, IMPACT, District of Columbia Public Schools, telephone interview, Oct. 14, 2014.

[168] Valerie Strauss, "Errors found in D.C. teacher evaluations (2nd update)," *Washington Post*, The Answer Sheet Blog, Dec. 23, 2013, http://www.washingtonpost.com/ (accessed May 6, 2014). Because of the relative nature of the value-added measure, equal numbers of teachers should have received higher ratings and lower ratings. Those teachers who should have received lower ratings were held harmless. Value-added scores were raised to correct the scores for teachers who should have received higher ratings. Alden Wells, Manager, IMPACT, and Luke Hostetter, Coordinator, IMPACT, District of Columbia Public Schools, e-mail attachment, Feb. 11, 2015.

[169] Susan Headden, *Inside IMPACT: D.C.'s Model Teacher Evaluation System*, Education Sector, 2011, p.11, http://www.educationsector.org/ (accessed June 20, 2014).

[170] District of Columbia Office of the Inspector General, *Report of Investigation into Cheating on the DC Comprehensive Assessment System Standardized Exams Administered by the District of Columbia Public Schools*, Aug. 8, 2012, pp. 11-12, http://dcps.dc.gov/ (accessed June 26, 2014); U.S. Department of Education Office of Inspector General, Closing Memorandum for File, 11-000491, Jan. 17, 2013, p.3, https://www2.ed.gov/ (accessed June 26, 2014).

Florida Profile

[171] Laws of Florida, Chapter 2011-1, "Student Success Act," http://laws.flrules.org/ (accessed Oct. 1, 2014).

[172] Florida Department of Education, "Educator Quality Legislation: Frequently Asked Questions," Nov. 2013, pp. 3-4, http://info.fldoe.org/ (accessed Oct. 3, 2014).

[173] *Florida Statutes*, 2014, 1012.34 (3)(a)1, http://www.leg.state.fl.us/ (accessed Oct. 2, 2014). School districts may request to use student achievement rather than student learning growth for the student performance component of teachers' evaluations for courses without statewide assessments or if it is demonstrated to be a more appropriate measure. *Florida Statutes* 1012.34 (7)(c).

[174] *Florida Statutes*, 2014, 1012.34 (3)(a)1.a, http://www.leg.state.fl.us/ (accessed Oct. 2, 2014). If a school district received an exemption under Florida's Race to the Top Memorandum of Understanding for Phase 2, it is allowed to weight student learning growth in teacher evaluations at 40 percent rather than 50 percent, subject to State Board approval after 2011-12.

[175] Jason Gaitanis, Policy Research and Accountability Coordinator, Florida Department of Education, telephone interview, Sept. 29, 2014; Eileen L. McDaniel, Chief, Educator Recruitment, Development, and Retention, Florida Department of Education, e-mail, Oct. 12, 2014.

[176] *Florida Statutes*, 2014, 1012.34 (8) and (9), http://www.leg.state.fl.us/ (accessed Oct. 2, 2014).

[177] *Florida Statutes*, 2014, 1012.34 (3)(a)1, http://www.leg.state.fl.us/ (accessed Oct. 2, 2014); Florida Statutes, 2014, 1008.22(6), http://www.leg.state.fl.us/ (accessed Oct. 2, 2014).

[178] *Florida Statutes*, 2014, 1012.34 (7)(c) and (d), http://www.leg.state.fl.us/ (accessed Oct. 2, 2014).

[179] *Florida Statutes*, 2014, 1012.34 (3)(a) 2 and 4, http://www.leg.state.fl.us/ (accessed Oct. 2, 2014).

[180] Student performance component may be reduced to 40 percent if three years of student data is not available for the teacher.

[181] Florida Department of Education, "Teacher Evaluation Instructional Practice Models," March 10, 2014, http://www.fldoe.org/ (accessed Dec. 2, 2014).

[182] Rules of Florida State Board of Education, 6A-5.065, the Educator Accomplished Practices, amended Feb. 13, 2011, https://www.flrules.org/ (accessed Dec. 2, 2014).

[183] Standards of Florida Department of Education, 6A-10.080, Code of Ethics of the Education Profession in Florida, amended July 6, 1982, and 6A-10.081, Principles of Professional Conduct for the Education Profession in Florida, amended Dec. 29, 1998, http://www.fldoe.org/ (accessed Oct. 2, 2014).

[184] *Florida Statutes*, 2014, 1012.34 (8), http://www.leg.state.fl.us/ (accessed Oct. 2, 2014).

[185] Laws of Florida, Chapter 2011-1, "Student Success Act," Section 4, pp. 12-13, http://laws.flrules.org/ (accessed Oct. 1, 2014).

[186] Florida Department of Education, "Educator Evaluations," presented at the Education Accountability Summit, Aug. 26-28, 2013, p. 14, http://www.fldoe.org/ (accessed Dec. 2, 2014).

[187] *Florida Statutes*, 2014, 1012.335 (2), http://www.leg.state.fl.us/ (accessed Oct. 3, 2014); Jason Gaitanis, Policy Research and Accountability Coordinator, Florida Department of Education, telephone interview, Sept. 29, 2014.

[188] Florida Statutes, 2014, 1012.33(1)(a) and 1012.33(4), http://www.leg.state.fl.us/ (accessed Oct. 3, 2014).

[189] Florida Department of Education, "Statewide Results of 2012-13 Evaluation Ratings," Dec. 3, 2013, p.1, http://www.fldoe.org/ (accessed Jan. 8, 2015).

[190] Florida Department of Education, "2013-14 Preliminary District Educator Evaluation Results," Nov. 19, 2014, p.3, http://www.fldoe.org/ (accessed Jan. 8, 2015).

[191] Laws of Florida, Chapter 2011-1, "Student Success Act," Section 4, part (e), and Section13, part (5), http://laws.flrules.org/ (accessed Oct. 1, 2014); Florida Department of Education, "Educator Quality Legislation: Frequently Asked Questions," Nov. 2013, p. 7, http://info.fldoe.org/ (accessed Oct. 3, 2014).

[192] American Institutes for Research, "Florida FCAT and Algebra I EOC 2013-14 Results," Aug. 2014, p.14, http://www.fldoe.org/ (accessed Oct. 1, 2014).

[193] American Institutes for Research, "Florida FCAT and Algebra I EOC 2013-14 Results," Aug. 2014, p.16, http://www.fldoe.org/ (accessed Oct. 1, 2014).

[194] *Recommendations of the Florida Student Growth Implementation Committee*, 2011, p. 6, http://www.fldoe.org/ (select Value-added Model While Paper) (accessed July 11, 2014).

[195] Jason Gaitanis, Policy, Research, and Accountability Coordinator, Florida Department of Education, e-mail, Oct. 13, 2014.

[196] American Institutes for Research, "Florida FCAT and Algebra I EOC 2013-14 Results," Aug. 2014, p. 8, http://www.fldoe.org/ (accessed Oct. 1, 2014).

[197] Ibid., p. 9.

[198] *Florida Statutes*, 2014, 1012.34, (7)(a), http://www.leg.state.fl.us/ (accessed Oct. 2, 2014).

[199] American Institutes for Research, "Florida FCAT and Algebra I EOC 2013-14 Results," Aug. 2014, pp.12-13, http://www.fldoe.org/ (accessed Oct. 1, 2014).

[200] Jason Gaitanis, Policy Research and Accountability Coordinator, Florida Department of Education, telephone interview, Sept. 29, 2014.

[201] American Institutes for Research, "Florida FCAT and Algebra I EOC 2013-14 Results," Aug. 2014, p. 5, http://www.fldoe.org/ (accessed Oct. 1, 2014).

[202] Jason Gaitanis, Policy, Research, and Accountability Coordinator, Florida Department of Education, e-mail, Oct. 13, 2014.

[203] Dr. Tony Bennett, Commissioner of Education, Florida Department of Education, "Request for Opinion Relating to the Value-added Model (VAM)," letter, addressed to Attorney General Bondi, Feb. 1, 2013, p. 3, footnote 13, https://www.scribd.com/ (accessed Dec. 1, 2014).

[204] Jason Gaitanis, Policy, Research, and Accountability Coordinator, Florida Department of Education, e-mail, Oct. 13, 2014.

[205] Topher Sanders, "Florida Times-Union sues state over teacher evaluation data," *The Florida Times Union – Jacksonville.com*, Feb. 7, 2013, http://jacksonville.com/ (accessed Dec. 1, 2014).

[206] *Florida Statutes*, 2014, 1012.31 (3)(a)2, http://www.flsenate.gov/ (accessed Oct. 3, 2014).

[207] District Court of Appeal, First District, State of Florida, Opinion in the case of Norris Publishing Group v. Florida Department of Education and the Florida Education Association, Nov. 12, 2013, http://opinions.1dca.org/ (accessed July 8, 2014).

[208] U.S. District Court for the Northern District of Florida, Gainesville Division, Order on Cross-Motions for Summary Judgment, Case No. 1:13-cv-72-MW-GRJ, Kim Cook et al. v. Florida Commissioner of Education, May 6, 2014, p.9, http://www.meyerbrookslaw.com/ (accessed Dec. 1, 2014). Because each district can make its own decisions about certain aspects of the evaluation system, the details vary somewhat among the teachers who filed suit based on which of the three districts they worked in.

[209] Ibid., pp.16-18.

[210] Valerie Strauss, "New Florida law: Teachers can't be evaluated on students they don't have," *Washington Post*, June 16, 2013, http://www.washingtonpost.com/ (accessed Oct 3, 2014); Laws of Florida, Chapter 2013-185, "An act relating to education, amending s. 1004.04, F.S., et al., Section 3," http://laws.flrules.org/ (accessed Dec. 2, 2014); Florida Department of Education, "Educator Quality Legislation: Frequently Asked Questions," Nov. 2013, p. 1, http://info.fldoe.org/ (accessed Oct. 3, 2014).

Los Angeles Unified School District Profile

[211] Doe et al v. Deasy et al, Tentative Decision on Permission for Writ of Mandate: Granted.

[212] Los Angeles Unified School District, Board of Education, AGT 101: Introduction to Academic Growth over Time, January 2011, pp. 41-43, http://achieve.lausd.net/ (accessed July 15, 2014).

[213] U.S. Department of Education, Teacher Incentive Fund, FY 2012 (Cohort 4) Awarded Grants, S374A120066 – Los Angeles Unified School District (CA), http://www2.ed.gov/ (accessed Oct. 2, 2014); Los Angeles Unified School District, Application to USDOE for Grants Under the TIF General Competition, Closing Date: July 27, 2012, p. 12, http://www2.ed.gov/ (accessed Oct. 2, 2014).

[214] U.S. Department of Education, Obama Administration Approves NCLB Waiver Request for California CORE Districts, Aug. 6, 2013, http://www.ed.gov/ (accessed Oct. 2, 2014).

[215] Los Angeles Unified School District and United Teachers Los Angeles, 2008-2011 Contract, (still in effect in 2014) http://www.utla.net/ (accessed Oct. 2, 2014); LAUSD-UTLA December 2012 Evaluation Procedures Supplement to Article X, pp. 1, 3, and 5, http://www.utla.net/ (accessed Oct. 2, 2014).

[216] Los Angeles Unified School District, Policy Bulletin 5335.1, Performance Evaluation for UTLA Bargaining Unit Personnel, Feb. 15, 2013, p. 1, http://www.teachinla.com/ (accessed July 16, 2014).

[217] Education Data Partnership, Negotiating Teachers' Contracts in California, Last updated Aug. 4, 2014, https://www.ed-data.k12.ca.us/ (accessed Oct. 2, 2014).

[218] Value-Added Research Center, Wisconsin Center for Education Research, University of Wisconsin-Madison, *Academic Growth over Time: Technical Report on the LAUSD Teacher-Level Model Academic Year 2012-2013*, Los Angeles Unified District, pp. 1-3, 24.

[219] Los Angeles Unified School District, Application to the U.S. Department of Education Under the TIF General Competition, CFDA # 84.374A, p. 29, http://www2.ed.gov/ (accessed Sept. 26, 2014).

[220] California Assembly Bill 484, http://legiscan.com/ (accessed Oct. 16, 2014); Rachel Bonkovsky, Executive Director, Talent Management, Los Angeles Unified School District, telephone interview, Oct. 14, 2014. The science tests in grades 5, 8, and 10 are continuing.

[221] Los Angeles Unified School District, Policy Bulletin 5335.1, Performance Evaluation for UTLA Bargaining Unit Personnel and Supplemental Agreement on Employee Evaluation Procedures: Frequently Asked Questions, Feb. 15, 2013, p. 2 (of the FAQs document), http://www.teachinla.com/ (accessed July 16, 2014); Rachel Bonkovsky, Executive Director, Talent Management, Los Angeles Unified School District, telephone interview, Oct. 14, 2014.

[222] LAUSD-UTLA December 2012 Evaluation Procedures Supplement to Article X, p. 3, http://www.utla.net/ (accessed Oct. 17, 2014).

[223] Los Angeles Unified School District, Policy Bulletin 5335.1, Performance Evaluation for UTLA Bargaining Unit Personnel, Feb. 15, 2013, pp. 6-7, http://www.teachinla.com/ (accessed July 16, 2014).

[224] Los Angeles Unified School District, Teacher Effectiveness Task Force, Final Report, April 13, 2010, p. 8, http://talentmanagement.lausd.net/ (accessed Oct. 1, 2014); Los Angeles Unified School District, LAUSD Teaching and Learning Framework, Talent Management Division, Released May 2014, p. 3, http://achieve.lausd.net/ (accessed Nov. 20, 2014); Los Angeles Unified School District, Information Brief, LAUSD's Teaching and Learning Framework, not dated, pp. 9-10, http://lausd.schoolwires.net/ (accessed Nov. 18, 2014).

[225] Los Angeles Unified School District, Support All Employees (SAE) Strategy and Academic Growth over Time (AGT), Frequently Asked Questions, Updated: November 2011, p. 29.

[226] Los Angeles Unified School District, Teacher Growth and Development Cycle, http://achieve.lausd.net/ (accessed Oct. 2, 2014).

[227] 2008-2011 Agreement, Los Angeles Unified School District and United Teachers Los Angeles, p. 65, http://www.utla.net/ (accessed July 17, 2014); National Council on Teacher Quality, Teacher Quality Roadmap: Improving Policies and Practice in LAUSD, not dated, p. 24, http://www.nctq.org/ (accessed July 17, 2014).

[228] Los Angeles Unified School District, Policy Bulletin 5335.1, Performance Evaluation for UTLA Bargaining Unit Personnel, Feb. 15, 2013, pp. 4-5, http://www.teachinla.com/ (accessed July 16, 2014).

[229] 2008-2011 Agreement, Los Angeles Unified School District and United Teachers Los Angeles, p. 65, http://www.utla.net/ (accessed July 17, 2014); National Council on Teacher Quality, Teacher Quality Roadmap: Improving Policies and Practice in LAUSD, not dated, p. 25, http://www.nctq.org/ (accessed July 17, 2014).

[230] Los Angeles Unified School District, Application for Grants under the TIF General Competition, Closing Date: July 27, 2012, pp. 14 and 36, https://www2.ed.gov/ (accessed Oct. 17, 2014).

[231] Rachel Bonkovsky, Executive Director, Talent Management, Los Angeles Unified School District, telephone interview, Oct. 14, 2014.

[232] LAUSD-UTLA December 2012 Evaluation Procedures Supplement to Article X, p. 5, http://www.utla.net/ (accessed Oct. 17, 2014).

[233] Los Angeles Unified School District, Academic Growth Over Time (AGT), Frequently Asked Questions, Updated: Sept. 2013, pp. 1 and 22.

[234] Ibid., pp. 25-26.

[235] Ibid., pp. 14-15.

[236] Ibid., p. 21.

[237] Value-Added Research Center, Wisconsin Center for Education Research, University of Wisconsin-Madison, Academic Growth over Time: Technical Report on the LAUSD Teacher-Level Model Academic Year 2012-2013, Los Angeles Unified District, p. 11.

[238] Ibid., pp. 18-19.

[239] Los Angeles Unified School District, Academic Growth Over Time (AGT), Frequently Asked Questions, Updated: Sept. 2013, p. 24.

[240] Los Angeles Times Communications LLC v. Los Angeles Unified School District, BS 139828, Tentative Decision, p. 26; Los Angeles Unified School District et al., Petitioners, v. The Superior Court of Los Angeles County, Respondent; Los Angeles Times Communications LLC, Real Party in Interest, B251693, Court of Appeal of California, Second Appellate District, Division Eight, July 23, 2014, Opinion Filed, p. 21, http://achieve.lausd.net/ (accessed Oct. 2, 2014).

North Carolina Profile

[241] Dayne Batten, Christopher Britt, Jennifer DeNeal, and Lauren Hales, *NC Teacher Evaluations and Teacher Effectiveness: Exploring the Relationship Between Value-Added Data and Teacher Evaluations*, Public Schools of North Carolina, Aug. 2012, pp. 1-2, http://www.ncpublicschools.org/ (accessed July 23, 2014). USDOE awarded North Carolina a $400 million grant in the Race to the Top competition in 2010. U.S. Department of Education, Nine States and the District of Columbia Win Second Round Race to the Top Grants, Aug. 24, 2010, http://www.ed.gov/ (Nov. 21, 2014).

[242] North Carolina, ESEA Flexibility Request, Revised May 24, 2012, http://www2.ed.gov/ (accessed Oct. 24, 2014); Arne Duncan, Secretary, U.S. Department of Education, letter to June St. Clair Atkinson, Superintendent, North Carolina Department of Public Instruction, May 29, 2014, http://www2.ed.gov/ (accessed Oct. 24, 2014).

[243] Deborah S. Delisle, Assistant Secretary, U.S. Department of Education, letter to June St. Clair Atkinson, Superintendent, North Carolina Department of Public Instruction, Aug. 14, 2014, http://www2.ed.gov/ (accessed Oct. 24, 2014).

[244] Deborah S. Delisle, Assistant Secretary, U.S. Department of Education, letter to Pat McCrory, Governor of North Carolina, and June St. Clair Atkinson, Superintendent, North Carolina Department of Public Instruction, Feb. 12, 2014, http://www2.ed.gov/ (accessed Nov. 21, 2014).

[245] Public Schools of North Carolina, Educator Effectiveness Model, Student Growth, http://www.ncpublicschools.org/ (accessed July 23, 2014).

[246] Heather Lynn, Nate Barrett, Julie Marks, Aubrey Comperatore, Gary T. Henry, and James Guthrie, *Measures of Student Growth in the North Carolina Educator Evaluation System*, Executive Summary, Sept. 2013, p. 3, http://cerenc.org/ (accessed July 23, 2014). Note that for various technical reasons EVAAS is not able to supply results for all of the exams in these categories. Public Schools of North Carolina, Value-Added Reporting on 2012-13 Assessments, http://www.ncpublicschools.org/ and Next Steps for NC Final Exams without Teacher-Level Value-Added Results, http://www.ncpublicschools.org/ (both accessed July 25, 2014).

[247] Southern Regional Education Board, State Implementation of Common Core State Standards: Evaluation of Teachers and Leaders, March 2014, p. 54, http://publications.sreb.org/ (accessed July 23, 2014); Leigh Kelley, "New teacher evaluation system being implemented," *BlueRidgeNow.com*, Feb. 10, 2013, http://www.blueridgenow.com/ (accessed July 23, 2014).

[248] North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 15, p. 6.

[249] Public Schools of North Carolina, News Releases 2012-13, North Carolina Enters End of Year Assessment Period, May 21, 2013, http://www.ncpublicschools.org/ (accessed Oct. 23, 2014); Ulrich Boser, *Race to the Top: What Have We Learned from the States So Far? A State-by-State Evaluation of Race to the Top Performance*, Center for American Progress, March 2012, p. 47, http://cdn.americanprogress.org/ (accessed July 24, 2014).

[250] Public Schools of North Carolina, Measuring Student Learning for Educator Effectiveness: A Guide to the Use of Student Growth Data in the Evaluation of North Carolina Teachers, Dec. 10, 2013, pp. 2-4, http://www.dpi.state.nc.us/ (accessed July 23, 2014).

[251] See the wiki developed by the North Carolina Department of Public Instruction, Analysis of Student Work (ASW), http://ncasw.ncdpi.wikispaces.net/.

[252] Public Schools of North Carolina, North Carolina Teacher Evaluation Process, originally published 2009, revised 2012, pp. 20-34, http://www.ncpublicschools.org/ (accessed July 23, 2014).

[253] In April 2013, the North Carolina State Board of Education adopted a policy that requires educators to receive an overall effectiveness status, which will incorporate ratings from all six teaching standards. Heather Lynn, Nate Barrett, Julie Marks, Aubrey Comperatore, Gary T. Henry, and James Guthrie, *Measures of Student Growth in the North Carolina Educator Evaluation System*, Executive Summary, Sept. 2013, p. 1, http://cerenc.org/ (accessed July 23, 2014).

[254] North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 8, p. 4.

[255] Heather Lynn, Nate Barrett, Julie Marks, Aubrey Comperatore, Gary T. Henry, and James Guthrie, *Measures of Student Growth in the North Carolina Educator Evaluation System*, Executive Summary, Sept. 2013, p. 1, http://cerenc.org/ (accessed July 23, 2014); Rebecca Garland, Chief Academic Officer, Public Schools of North Carolina, Memorandum on State Board of Education Actions Regarding Educator Effectiveness, To Superintendents, Associate Superintendents for Curriculum and Instruction, Human Resource Directors, Testing and Accountability Directors/ Coordinators, and Charter School Directors, Oct. 10, 2013,p. 3, http://www.dpi.state.nc.us/ (accessed Oct. 21, 2014).

[256] Deborah S. Delisle, Assistant Secretary, and Ann Whalen, Director, Policy and Program Implementation, Implementation and Support Unit, U.S. Department of Education, Letter to Pat McCrory, Governor, and June St. Clair Arkinson, Superintendent of Public Instruction, North Carolina, Feb. 12, 2014, http://www2.ed.gov/ (accessed July 23, 2014).

[257] North Carolina General Statutes 115C-333.1, (a) (Effective July 1, 2014, until June 30, 2018) and (a) (Effective June 30, 2018), http://www.ncga.state.nc.us/ (accessed Oct. 24, 2014).

[258] Rebecca Garland, Chief Academic Officer, Public Schools of North Carolina, Memorandum on State Board of Education Actions Regarding Educator Effectiveness, To Superintendents, Associate Superintendents for Curriculum and Instruction, Human Resource Directors, Testing and Accountability Directors/ Coordinators, and Charter School Directors, Oct. 10, 2013, http://www.dpi.state.nc.us/ (accessed Oct. 21, 2014).

[259] Public Schools of North Carolina, North Carolina Teacher Evaluation Process, originally published 2009, revised 2012, p. 18, http://www.ncpublicschools.org/ (accessed July 23, 2014); North Carolina General Statutes 115C-333.1, (b), (d), and (f) (Effective July 1, 2014, until June 30, 2018) and (Effective June 30, 2018), http://www.ncga.state.nc.us/ (accessed Oct. 24, 2014).

[260] Career Pathways for Teachers: A Framework for Student and Teacher Success, not dated, pp. 2 and 4, http://www.ncpapa.org/ (accessed Oct. 24, 2014).

[261] North Carolina State Board of Education, Minutes, Feb. 29, 2012, p. 12, www.ncpublicschools.org/ (accessed July 24, 2014).

[262] SAS EVAAS for K-12, North Carolina: Resource Guide for Value-Added Reporting, 2007, http://www.dpi.state.nc.us/ (accessed July 24, 2014); Ulrich Boser, *Race to the Top: What Have We Learned from the States So Far? A State-by-State Evaluation of Race to the Top Performance*, Center for American Progress, March 2012, p. 47, http://cdn.americanprogress.org/ (accessed July 24, 2014).

[263] North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 8, p. 4.

[264] SAS EVAAS, How does EVAAS accommodate changes in testing regimes?, June 2013, pp. 1-2, http://www.ncpublicschools.org/ (accessed Nov. 24, 2014); North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, p. 10.

[265] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, March 25, 2010, p. 9, http://www.sas.com/ (accessed July 24, 2014).

[266] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, March 25, 2010, pp. 3 and 9, http://www.sas.com/ (accessed July 24, 2014).

[267] North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 29, p. 11.

[268] SAS, Ohio Department of Education: Technical Documentation of EVAAS Analyses, June 10, 2014, p. 6, http://education.ohio.gov/ (accessed Nov. 24, 2014).

[269] S. Paul Wright, John T. White, William L. Sanders, and June C. Rivers, *SAS EVAAS Statistical Models*, March 25, 2010, p. 9, http://www.sas.com/ (accessed July 24, 2014).

[270] North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 8, p. 4.

[271] North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 11, p. 5.

[272] SAS Institute, Interpreting the Teacher Value-Added Report, North Carolina Department of Public Instruction, 2012, p. 1, http://si2013.ncdpi.wikispaces.net/ (accessed Nov. 24, 2014).

[273] Public Schools of North Carolina, PowerPoint Presentation: Exploring EVAAS in a Common Core World, p. 50, (accessed July 24, 2014). North Carolina Department of Public instruction, Frequently Asked Questions on EVAAS, Spring 2013, Question 23, p. 8.

[274] Heather Lynn, Nate Barrett, Julie Marks, Aubrey Comperatore, Gary T. Henry, and James Guthrie, *Measures of Student Growth in the North Carolina Educator Evaluation System*, Executive Summary, Consortium for Educational Research and Evaluation—North Carolina, Sept. 2013, p. 5, http://cerenc.org/ (accessed July 23, 2014).

Ohio Profile

[275] KidsOhio.org, "Ohio's Race to the Top Effort Continues amid Leadership Changes," Summer 2011, http://www.kidsohio.org/ (accessed Nov. 25, 2014).

[276] Ohio Department of Education, "Business Rules for Student Growth Measures," Nov. 2013. Teachers who only taught some classes with value-added data would have the value-added component reduced proportionately to their teaching assignments, as long as it was weighted as at least 10 percent.

[277] Patrick O'Donnell, "Change in how teachers are graded in Ohio is a compromise in a big debate," *The Plain Dealer*, http://www.cleveland.com/ (accessed June 13, 2014).

[278] Patrick O'Donnell, "Ohio teachers will be graded on students' academic growth," *The Plain Dealer*, Nov. 13, 2011, http://blog.cleveland.com/ (accessed Nov. 26, 2014).

[279] Ohio Department of Education, *Ohio Teacher Evaluation System, 2014 Model*, Sept. 26, 2014, p. 6, http://education.ohio.gov/ (accessed Oct. 29, 2014).

[280] Ibid.

[281] Ohio Department of Education, "FAQs about HB 362 and Teacher Evaluation," p.1, [no date], http://education.ohio.gov/ (accessed Oct. 31, 2014).

[282] Thomas B. Fordham Institute, *Ohio Value-added Primer*, August 25, 2008, p.3, http://edexcellence.net/ (accessed May 6, 2014).

[283] SAS Institute, Inc., *Technical Documentation of EVAAS Analyses*, Ohio Department of Education, June 10, 2014, p. 6, http://education.ohio.gov/ (accessed Nov. 4, 2014).

[284] SAS Institute, Inc., *Technical Documentation of EVAAS Analyses*, Ohio Department of Education, June 10, 2014, pp. 13-14, http://education.ohio.gov/ (accessed Nov. 4, 2014).

[285] SAS Institute, Inc., *Technical Documentation of EVAAS Analyses*, Ohio Department of Education, June 10, 2014, p. 9, http://education.ohio.gov/ (accessed Nov. 4, 2014). When Ohio switches from the Ohio Achievement Assessments to its new state tests in 2014-15, it is expected that an intra-year benchmark will be used instead of the current base year benchmark, as is necessary when there is a change in the standardized testing programs and the new test is not on the same scale as the old test.

[286] SAS Institute, Inc., *Technical Documentation of EVAAS Analyses*, Ohio Department of Education, June 10, 2014, p. 20, http://education.ohio.gov/ (accessed Nov. 4, 2014).

[287] Ibid, pp. 28-29.

[288] Patrick O'Donnell, "Teachers' 'value-added' ratings and relationship to student income levels questioned," *The Plain Dealer,* June 18, 2013, http://impact.cleveland.com/ (accessed Nov. 26, 2014).

[289] Julia Simmerer, Senior Executive Director, Center for the Teaching Profession, Ohio Department of Education (by T. Mumm), "RE: Questions on Ohio's Teacher Evaluation System," e-mail, Nov. 6, 2014.

[290] Julia Simmerer, Senior Executive Director, Center for the Teaching Profession, Ohio Department of Education (by T. Mumm), "RE: Questions on Ohio's Teacher Evaluation System," e-mail, Nov. 6, 2014.

Offices of Research and Education Accountability
Phillip Doss, Director
Suite 1700, James K. Polk Building ▪ 505 Deaderick Street
Nashville, Tennessee 37243 ▪ (615) 401-7866
www.comptroller.tn.gov/orea