



An examination of performance-based teacher evaluation systems in five states





An examination of performance-based teacher evaluation systems in five states

February 2012

Prepared by

Karen Shakman
Education Development Center, Inc.

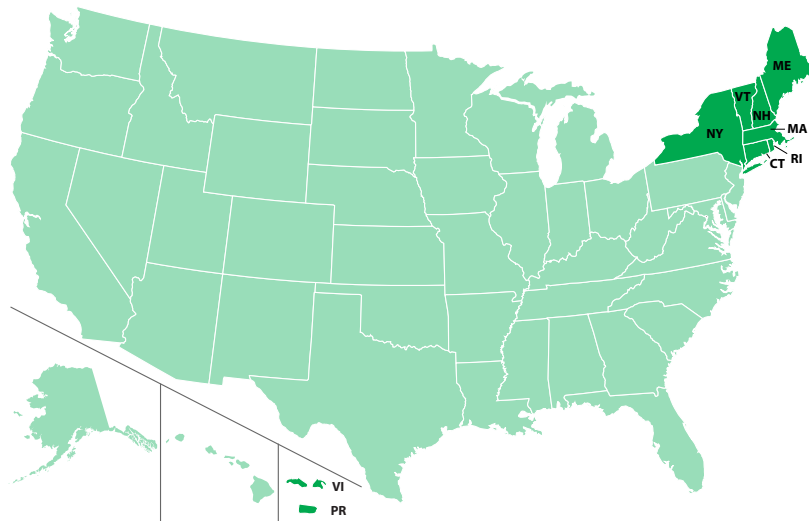
Julie Riordan
Education Development Center, Inc.

María Teresa Sánchez
Education Development Center, Inc.

Kyle DeMeo Cook
Education Development Center, Inc.

Richard Fournier
Education Development Center, Inc.

Jessica Brett
Education Development Center, Inc.



Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

February 2012

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0025 by Regional Educational Laboratory Northeast and Islands administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Shakman, K., Riordan, J, Sánchez, M.T., DeMeo Cook, K., Fournier, R., and Brett, J. (2012). *An examination of performance-based teacher evaluation systems in five states*. (Issues & Answers Report, REL 2012–No. 129). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.

An examination of performance-based teacher evaluation systems in five states

This study of performance-based teacher evaluation systems in the five states that had implemented statewide systems as of 2010/11 finds considerable variation among them. However, all five states' systems include observations, self-assessments, and multiple rating categories. In addition, the evaluation rubrics in each state reflect most of the teaching standards set out by the Interstate Teacher Assessment and Support Consortium.

A combination of research and federal and state interest in measuring teacher effectiveness has galvanized support for reform of teacher evaluation systems. A number of researchers have called for multiple measures of teacher effectiveness, greater differentiation among teachers, and stronger connections to outcomes for students (Toch and Rothman 2008; Gordon, Kane, and Staiger 2006; Hene-man et al. 2006). The application guidelines for the 2009 Race to the Top federal grant competition called for states to develop systems that evaluate teacher effectiveness using multiple rating categories, not the traditional binary rating of satisfactory to unsatisfactory, and to take into account data on student growth (U.S. Department of Education 2009). In response to this new policy direction, many states' Race to the Top grant proposals provided plans for changes to their teacher evaluation systems.

This study reports on performance-based teacher evaluation systems in five states that have implemented such systems. It investigates two primary research questions:

- What are the key characteristics of state-level performance-based teacher evaluation systems in the study states?
- How do state teacher evaluation measures, the teaching standards the evaluations are designed to measure, and rating categories differ across states that have implemented statewide systems?

To answer these questions, the study team reviewed state education agency websites and publicly available documents for all 50 states to identify states whose performance-based teacher evaluation systems met the following criteria:

- Was required for practicing general educators.
- Was operational statewide as of the 2010/11 school year.
- Included multiple rating categories.
- Used multiple measures of teacher effectiveness, such as observations, self-assessments, and professional growth plans.

Five states (Delaware, Georgia, North Carolina, Tennessee, and Texas) met these criteria. Key study findings include the following:

- Of the five states that met the criteria, three have new systems (1–3 years old), and two have systems that are more than 10 years old.
- One state (Georgia) requires full annual evaluations for all teachers. In the other states, evaluations are annual for teachers whom the state defines as novice and less frequent or less comprehensive for more experienced teachers.
- All five states include self-assessments and observations of classroom teaching as part of teacher assessment. States differ in who conducts the observations, how often evaluations are conducted, and what scoring parameters are used.
- In each of the five states, teacher evaluation rubrics and scoring forms reflect most or all of the 10 teaching standards set forth by the Interstate Teacher Assessment and Support Consortium (InTASC). These standards relate to teachers' understanding of how students learn, content knowledge, instructional practice, and professional responsibilities. All 10 standards are reflected in the teacher evaluation rubrics in North Carolina and Texas, 9 are reflected in Georgia, and 8 are reflected in Delaware and Tennessee. One InTASC standard—specifying that teachers demonstrate an understanding of how students learn—is absent in two states' evaluation rubrics (Georgia and Tennessee).
- States differ in the number of rating categories used and how they compute scores and determine passing scores.

February 2012

TABLE OF CONTENTS
Why this study? 1

- National interest in increasing teacher effectiveness 1
- Regional need for information on teacher evaluation 2
- Defining performance-based teacher evaluation 3
- Research questions 3

Study findings 4

- Key characteristics of state-level performance-based teacher evaluation systems 4
- Teacher evaluation measures, teaching standards, and rating categories 7

Study limitations 13**Directions for future research 13****Notes 16****Appendix A State profiles 17****Appendix B Methodology 27****References 31****Boxes**

- 1 Key terms 2
- 2 Data sources and methods 5
- B1 Text of email sent to state officials requesting information statewide performance-based teacher evaluation systems 29

Tables

- 1 Measures included in state-level performance-based teacher evaluation systems in five states 8
- 2 Teaching domains and standards in performance-based teacher evaluation systems in study states 10
- 3 Interstate Teacher Assessment and Support Consortium teaching standards incorporated into evaluation rubrics in study states 12
- 4 Rating categories and methods for calculating overall teacher rating in study states 14
- A1 Profile of Delaware's Performance-Based Teacher Evaluation System 17
- A2 Links to key resources on Delaware's Performance-Based Teacher Evaluation System 18
- A3 Profile of Georgia's Performance-Based Teacher Evaluation System 19
- A4 Links to key resources on Georgia's Performance-Based Teacher Evaluation System 20
- A5 Profile of North Carolina's Performance-Based Teacher Evaluation System 21
- A6 Links to key resources on North Carolina's Performance-Based Teacher Evaluation System 22
- A7 Profile of Tennessee's Performance-Based Teacher Evaluation System 23

A8	Links to key resources on Tennessee’s Performance-Based Teacher Evaluation System	24
A9	Profile of Texas’s Performance-Based Teacher Evaluation System	24
A10	Links to key resources on the Texas’s Professional Development and Appraisal System	26
B1	Terms used to search for state-level performance-based teacher evaluation systems	27
B2	Form used to record information on states’ performance-based teacher evaluation systems	27
B3	Overview of performance-based teacher evaluation systems in study states	28
B4	Interrater reliability of coding of state performance-based teacher evaluation systems	30

This study of performance-based teacher evaluation systems in the five states that had implemented statewide systems as of 2010/11 finds considerable variation among them. However, all five states' systems include observations, self-assessments, and multiple rating categories. In addition, the evaluation rubrics in each state reflect most of the teaching standards set out by the Interstate Teacher Assessment and Support Consortium.

WHY THIS STUDY?

Recent studies have highlighted the weak state of teacher evaluation and the need for reform (Measures of Effective Teaching Project 2010; National Council on Teacher Quality 2009; Weisberg et al. 2009; Toch and Rothman 2008). Most teacher evaluations neither differentiate among teachers and the quality of their instruction nor emphasize teachers' influence on student achievement (Daley and Kim 2010; Measures of Effective Teaching Project 2010; Weisberg et al. 2009). The widespread use of binary rating systems, in which teachers receive an overall rating of either satisfactory or unsatisfactory, has been criticized for lacking rigor, as nearly 99 percent of teachers in some districts earn satisfactory ratings (Weisberg et al. 2009). Formal teaching qualifications (such as degrees and certification), which sometimes are used to evaluate and reward teachers, are weakly correlated with student achievement (Toch and Rothman 2008; Aaronson, Barrow, and Sander 2007; Kane, Rockoff, and Staiger 2006). Additionally, research indicates that principals can generally identify teachers who are the most and least effective but are less able to differentiate among teachers whose effectiveness is between these extremes (Jacobs and Lefgren 2008).

National interest in increasing teacher effectiveness

Interest in educator effectiveness, specifically in teacher evaluation, has grown in recent years, partly in response to the emphasis on effective teachers that is evident in Race to the Top, the competitive federal grant awards program. The Race to the Top guidelines for state teacher evaluation systems call for states to develop "rigorous, transparent, and fair evaluation systems . . . that . . . differentiate effectiveness using multiple rating categories that take into account data on student growth . . . as a significant factor" (U.S. Department of Education 2009, p. 9). In response to these guidelines, states across the country proposed major reforms that would create comprehensive evaluation systems with multiple measures of teacher performance, including measures

of student growth, observations of teachers, analysis of teacher artifacts (such as lesson plans, assessments, assignments, rubrics, student work, or portfolios), peer review, student reflections and feedback, and participation in professional development (Learning Point Associates 2010). (For a definition of multiple measures and other key terms used in this report, see box 1).

Regional need for information on teacher evaluation

The Regional Educational Laboratory (REL) Northeast and Islands received several requests from schools, districts, and state education agencies for more information on educator evaluations. In 2010, REL Northeast and Islands completed a technical assistance project for the New York State Education

Department's Associate Commissioner for Higher Education that examined performance assessments linked to both initial and continuing certification. This project elicited considerable interest from stakeholders in the region, who requested information about the kinds of evaluation systems other states have in place or are in the process of implementing. At the REL Northeast and Islands governing board meetings in 2010 and 2011, members inquired about effective models of teacher evaluation, the use of student achievement data in teacher evaluation, and the role of administrators in supporting and managing evaluation systems.

The three Race to the Top states in the Northeast and Islands Region (Massachusetts, New York, and Rhode Island) are developing and implementing

BOX 1

Key terms

Domains. General bodies of knowledge and skills for teaching.

Evaluation measures. The specific tools and approaches, such as classroom observation, analysis of classroom artifacts, and portfolios, used to support the measurement of teacher effectiveness.

Multiple measures. Multiple indicators that target a range of components of effective teaching, using such data sources as classroom observations, pre- and post-conference, self-assessments, analysis of classroom artifacts, and professional growth plans.

Multiple rating categories. The categories or terms that differentiate teacher proficiency across three or more levels, such as unsatisfactory, meets expectations, above expectations, and exemplary.

Race to the Top. The \$4.35 billion competitive grant program designed to encourage and reward states that have demonstrated success in raising student achievement and that have developed strong plans to accelerate their reforms in the future.

Rubrics or scoring forms. The evaluation forms, with different levels of proficiency described across the multiple rating categories, used to rate or score teacher performance according to the teaching standards the evaluation system is designed to measure. Not all states use a rubric, but all states included in this study have some type of summative form for rating or scoring teacher performance.

Student growth data. Data used to measure “a change in student achievement for an individual student between two or more points in time” (U. S. Department of Education 2009, p. 9). Approaches that use student growth data to measure teacher performance are sometimes referred

to as “value-added” approaches or models.

Teaching or classroom artifacts. Lesson plans, assessments, assignments, rubrics, or student work that may be used as evidence of teachers’ professional practice.

Teaching standards. The knowledge and skills teachers should possess. States use various terms to refer to the main bodies of knowledge and skills for teaching, including *domains*, *strands*, and *standards*. They refer to the more specific types of knowledge and skills that teachers should be able to demonstrate as *standards*, *criteria*, *elements*, or *indicators*. In this report, the term *domains* refers to very general bodies of knowledge (such as planning or instruction); the term *standards* refers to more specific types of knowledge and skills that teachers should demonstrate and according to which they are evaluated. States’ own terminology is used in the profiles provided in appendix A.

new systems of evaluation. Massachusetts established a task force charged with developing a framework for evaluation, and the Massachusetts Board of Elementary and Secondary Education recently passed new teacher evaluation regulations based on the task force proposal. New York completed a similar process and passed a state law requiring a new approach to performance evaluation. Rhode Island is establishing a new system, with three potential systems currently in development. In all three states, implementation of pilot evaluation systems is planned for the 2011/12 school year. Two other states in the region, Maine and New Hampshire, also have begun statewide efforts to reform their evaluation systems.

The New England Collaborative for Educator Effectiveness, a group of state education leaders from the six New England states, has been meeting since July 2009 to learn from research, experts, and each other how to develop new systems to evaluate educator effectiveness. Group members have identified the development of multiple measures of teacher effectiveness as a key priority and asked REL Northeast and Islands for support in learning about models that are already in place.

Defining performance-based teacher evaluation

A performance-based teacher evaluation system includes multiple measures of teacher performance and provides a range of evidence, demonstrating teacher knowledge and skills, related particularly to student achievement. Goe, Bell, and Little's (2008) review of 120 studies on measuring teacher effectiveness describes three different but related types of measures:

- Inputs, such as certification and licensure, content knowledge, and educational attainment.
- Processes, such as interactions among teachers and students in the classroom and interactions among teachers.
- Outputs, such as influence on student achievement and graduation rates.

Goe, Bell, and Little (2008) conclude that the use of multiple measures built on the elements of inputs, processes, and outputs is critical to capturing the range of knowledge and skills that make a teacher effective. They argue that a comprehensive assessment of teacher effectiveness should address multiple components of teacher effectiveness.

A performance-based teacher evaluation system includes multiple measures of teacher performance and provides a range of evidence, demonstrating teacher knowledge and skills, related particularly to student achievement

Cogshall, Max, and Bassett (2008) define performance-based assessment as a set of measurements of different aspects of teaching using multiple sources of evidence that provide both formative and summative feedback. Sources of evidence include classroom observation protocols, teacher-developed portfolios, lesson plans, sample individualized education programs for teachers, teacher responses to real-world teaching scenarios, and video records of instructional practice.

Among the multiple measures that may make up a performance-based evaluation system, measures of student performance have received considerable attention. Because student growth measures provide information about how teachers may affect student achievement, a broad range of scholars, including measurement and evaluation experts, high-stakes testing experts, and value-added model scholars, support the use of student growth measures as one of several performance measures (Daley and Kim 2010; Milanowski, Heneman, and Kimball 2009; Braun 2005).

Research questions

This project uses information from publicly available documents to answer two research questions:

- What are the key characteristics of state-level performance-based teacher evaluation systems in the study states?

This study provides an overview of five states' performance-based teacher evaluation systems and assesses their similarities and differences

- How do state teacher evaluation measures, the teaching standards the evaluations are designed to measure, and rating categories differ across states that have implemented statewide systems?

To answer these questions, the study team reviewed state education agency websites and publicly available documents for all 50 states to identify states whose performance-based teacher evaluation systems met the following criteria:

- Was required for practicing general educators.
- Was operational statewide as of the 2010/11 school year.
- Included multiple rating categories.
- Used multiple measures of teacher effectiveness.

Race to the Top guidelines for performance-based evaluation also call for systems to include student growth data as a “significant”¹ factor and to require annual evaluations of all teachers. Because no states included student growth data as a significant factor in their teacher evaluation criterion in 2010/11 (individual districts may have used such measures), the use of student growth data was not identified as a selection criterion.

Race to the Top guidelines also require annual evaluations of all teachers, but not all the states examined in this report conduct annual performance-based evaluations for all teachers. The states' evaluation timelines vary based on whether the teachers are novice or experienced.

The selection criteria were established to ensure inclusion only of:

- States with systems for evaluating the majority of practicing teachers. Evaluations designed for specific teaching populations,

principals, or other administrators are not the subject of this study.

- States in which systems were being implemented (rather than planned reforms), in order to provide information on how these systems are structured.
- States whose system met some of the basic requirements of a strong performance-based educator evaluation system identified in the research literature and the Race to the Top guidelines. (For a summary of the selection criterion, see table B3 in appendix B.)

Five states met the selection criteria. Four of the five (Delaware, Georgia, North Carolina, and Tennessee) are Race to the Top winners; the fifth state, Texas, did not apply for the competitive grant. This study provides an overview of these states' performance-based teacher evaluation systems and assesses their similarities and differences.

Between March and May 2011, the study team systematically collected data on these states and constructed profiles on their teacher evaluation systems. (See appendix A for the state profiles; see box 2 and appendix B for more detail on data sources and study methodology.)

STUDY FINDINGS

This section describes the key characteristics of the teacher evaluation systems in the five study states. It shows how the systems differ in the measures used to evaluate teachers, the teaching standards the evaluations are designed to measure, and the categories used to rate teachers.

Key characteristics of state-level performance-based teacher evaluation systems

Of the five states that met the study criteria, three have new systems (1–3 years old), and two have systems that are more than 10 years old. Both of the states with older systems have made changes

BOX 2

Data sources and methods

The study team obtained information on performance-based teacher evaluation systems for each of the five states in the study from the state agencies' web pages as well as from general Internet searches. It also emailed the five state education agencies asking for additional publicly available information on the state's teacher evaluation systems (this effort yielded no additional information).

Data sources. The following data sources were used:

- *General and other web pages of state education agencies.* General information about the states' evaluation systems was available on the state education agency web pages, which provide basic information about the overall system structure, measures used, and project timeline.
- *Guides and manuals.* Each state provides evaluation guides or manuals as publicly available downloadable resources. These manuals provide detail about the evaluation process, the measures

used, the frequency of evaluations, the standards by which teachers are evaluated, and the rubrics (or scoring forms) used. Some manuals also include historical information and details about state regulations for teacher evaluation.

- *Evaluation rubrics and forms.* Each state has developed evaluation forms for rating or scoring teacher performance according to the teaching standards the evaluation system is designed to measure. Some states use a traditional rubric, which includes various levels of proficiency across the multiple rating categories; all states have some type of summative scoring form for rating or scoring teachers' performance, however.
- *Regulations and legislation.* The authorizing legislation or regulations for each evaluation system provide additional information about the requirements and history of each evaluation system.
- *Program reports.* Only Delaware provides program reports about its evaluation system on the state

education website. These reports provide information about internal evaluations of the system, which has been in place since 2005/06, the first pilot year of the Delaware Performance Assessment System II system.

- *Other documents.* Each state provides slightly different information and houses the information in different places. A general category ("other documents") reflects this range of material.

Study sample and analysis. The study team scanned all 50 states to identify states that met the study criteria. It then constructed profiles on the five states that met the criteria, based on the information available on each state's website. The study team used the teaching standards of the Interstate Teacher Assessment and Support Consortium as a priori codes in order to compare the teaching standards the evaluations are designed to measure across states. Two study team members independently coded the information, with unresolved discrepancies reconciled by a third team member. (For a full discussion of the study's methodology, see appendix A.)

several times since initial implementation. Only Georgia requires full annual evaluations for all teachers. The other states require annual evaluations for early career teachers and less frequent or less comprehensive evaluations for more experienced teachers. The only exception is Texas, which allows districts some freedom to determine the frequency of evaluation for all teachers.

All five states include observations and self-assessments as part of teacher assessment. States

differ in who conducts the observation, how often evaluations are conducted, and what scoring parameters are used.

Teacher evaluation rubrics and scoring forms in the five states reflect most or all of the teaching standards set forth by the Interstate Teacher Assessment and Support Consortium (InTASC), an organization formed by the Council of Chief State School Officers. The 10 InTASC standards relate to students and how they learn, teachers'

All five states include observations and self-assessments as part of teacher assessment, but states differ in who conducts the observation, how often evaluations are conducted, and what scoring parameters are used

content knowledge, instructional practices, and professional responsibilities. All 10 standards are reflected in the teacher evaluation rubrics in North Carolina and Texas, 9 are reflected in Georgia, and 8 are reflected in Delaware and Tennessee. One InTASC standard—specifying that teachers demonstrate an understanding of how students learn—is absent in two states’ evaluation rubrics (Georgia and Tennessee).

All five states use multiple rating categories. Systems vary, however, in how many rating categories the evaluations include, how scores are computed, and how a passing score is determined.

All states’ evaluation systems seek to facilitate the professional growth of teachers and assess the quality of teacher performance. Three states (Georgia, Tennessee, and Texas) make explicit reference to student learning in the stated goals of the evaluation systems. Tennessee is the only state that explicitly references a link between teacher evaluation and student growth.

The following sections describe each state’s teacher evaluation system. Full profiles of each state’s system appear in appendix A.

Delaware. The Delaware Performance Assessment System II (DPAS II) has been in place since the 2008/09 school year. Its stated purpose is to facilitate professional growth and continuous improvement of teachers and to serve as an instrument of quality assurance. Delaware teachers are evaluated on five domains: planning and preparation, classroom environment, instruction, professional responsibilities, and student improvement. Novice teachers are evaluated twice a year and experienced teachers once a year, based on observations, conferences, and a teacher self-assessment.

Georgia. The Georgia Classroom Analysis of State Standards Keys Teacher Evaluation System (CLASS

Keys) was established in 2010, in order to foster the individual professional growth (continuous improvement) of teachers and document teacher performance and quality. CLASS Keys includes five domains: curriculum and planning, standards-based instruction, assessment of student learning, professionalism, and student achievement. Teachers are evaluated annually, in three phases: teachers self-assess their level of performance and develop a draft of their professional growth plan; evaluators observe classrooms and collect additional evidence through conferences, meetings, and examination of student and teacher products; and evaluators rate teachers’ performance on each of the standards/ domains by reviewing all of the evidence collected during the year.

North Carolina. The North Carolina Teacher Evaluation Process (NCTEP) was introduced statewide in the 2010/11 school year. Its purpose is to assess teacher performance in relation to the North Carolina Professional Teaching Standards and to develop the growth of practitioners. The evaluation is based on five domains, which require that teachers demonstrate leadership, establish a respectful environment for a diverse population of students, know the content they teach, facilitate learning for their students, and reflect on their practice. The NCTEP includes eight components: training, orientation, teacher self-assessment, a pre-observation conference, observations, a post-observation conference, a summary evaluation conference and scoring of the teacher summary rating form, and professional development plans. Teachers who have reached “career status” (that is, have tenure) are evaluated at least once every five years; all other teachers are evaluated annually. Regardless of status, all teachers must participate in orientation, self-assessment, and professional development planning every year.

Tennessee. Tennessee’s teacher evaluation system, the Framework for Evaluation and Professional Growth, was introduced statewide in July 2000 and revised in 2004 and 2009. The system’s stated purpose is to encourage teachers to move beyond their level of performance by focusing on

student growth, self-reflection on areas for their own growth, and school improvement. Tennessee teachers are assessed in six domains: planning, teaching strategies, assessment and evaluation, learning environment, professional growth, and communication. The system includes four components: a teacher self-assessment, in which the teacher is asked to reflect on areas of strength and opportunities for growth; an educator's information record, in which teachers provide specific examples of analysis of student assessment data and professional growth activities; observations; and a future growth plan. Teachers also have the option of submitting a sample unit and lesson plan for review by the evaluator. Tenured teachers are evaluated every five years, with a minimum of two observations; novice teachers are evaluated annually, with a minimum of three observations a year in the first two years of teaching and two observations in the third year.

Texas. The Professional Development Appraisal System (PDAS) was developed in 1995 and first implemented statewide in 1997/98, with the purpose of improving student learning through the professional development of educators. The appraisal is based on eight domains: active, successful student participation in the learning process; learner-centered instruction; evaluation and feedback on student progress; management of student discipline, instructional strategies, time, and materials; professional communication; professional development; compliance with policies, operating procedures, and requirements; and improvement of academic performance of all students on the campus. Competency in each domain is measured by classroom observations and walkthroughs, teacher self-reports, and student performance. Teachers are evaluated every year, although districts may exempt qualified teachers from annual appraisals as long as they appraise them once every five years.

Teacher evaluation measures, teaching standards, and rating categories

Evaluation measures. Based on their review of 120 articles, Goe, Bell, and Little (2008) categorize

instruments that are either in use or represent “promising measures of teaching” that directly assess teachers’ classroom processes and activities. These measures include classroom observations, evaluations by principals, analysis of classroom artifacts, analysis of teaching portfolios, teacher self-reports of practice, student ratings of teacher performance, and value-added (student growth) strategies (described later in this report). The study team used these categories as a guide to organize the evaluation measures used, modifying one category (“principal evaluations” became “administrator evaluation”) and adding another (professional development/growth plans).

All five study states use multiple measures to evaluate teachers, including observation, an evaluation by an administrator, and some type of teacher self-assessment (table 1). All states except Delaware include some type of professional growth plan as a component of the evaluation. One state (Georgia) requires analysis of classroom artifacts; Tennessee includes analysis of classroom artifacts as an optional part of the evaluation.

States are also similar in the types of measures that are absent from their systems. No study states use student ratings of teachers or student growth data as measures of teacher performance. The only other measure identified by Goe, Bell, and Little (2008) that was not used by any of the five study states is analysis of teaching portfolios.

Classroom observation. All five states require classroom observations, although they vary in the number, length, and nature of the observations (for example, whether these observations are announced or unannounced and who conducts them). The number of observations required ranges from one to four a year. For example, in Delaware novice teachers (teachers holding an initial license) are observed at least twice a year, including

All five study states use multiple measures to evaluate teachers, including observation, an evaluation by an administrator, and some type of teacher self-assessment

TABLE 1

Measures included in state-level performance-based teacher evaluation systems in five states

State	Classroom observation	Administrator evaluation	Analysis of classroom artifacts	Analysis of teaching portfolios	Self-report of teacher practice	Student ratings of teacher performance	Value-added models	Professional development/growth plans
Delaware	✓	✓			✓			
Georgia	✓	✓	✓		✓			✓
North Carolina	✓	✓			✓			✓
Tennessee	✓	✓	a		✓			✓
Texas	✓	✓			✓			

a. An optional form analyzes the lesson plan.

Source: Authors' analysis of publicly available state documents; see appendix A for details and for more information on each system.

at least one announced and one unannounced observation. North Carolina requires that principals conduct at least three planned observations of probationary teachers (teachers without tenure and within their first four years of teaching), with each observation lasting at least 45 minutes, and that peer evaluators conduct at least one additional planned observation. These observations include a pre-conference and a post-conference attended by the teacher and observer.

Administrator evaluation. All five states require an evaluation by an administrator. In some states, the evaluator is the school principal; in other states, the evaluator can be a supervisor or district-level administrator.² Some states include an evaluation by an administrator as a discrete component of the evaluation system; in other states, the administrator conducts all or most of the evaluation but does not complete a separate administrator evaluation as a discrete component of the system.

Analysis of classroom artifacts. Goe, Bell, and Little (2008) define analysis of classroom or instructional artifacts as the use of a structured protocol to analyze artifacts such as lesson plans, assessments, assignments, rubrics, and student work. All five states assess classroom artifacts in some capacity, although not all include this as an explicit and standalone component, nor do most of the states have structured protocols for

analyzing artifacts. In Georgia, evaluators collect artifacts (such as lesson plans, data records, and assessments) that are related to the five required domains and assign a score for each domain. In other states, such as North Carolina, a lesson plan is required as a part of the classroom observation, but it is not scored directly. What sets Georgia apart from the other states is that these artifacts are discrete components of the evaluation system and receive their own score. In Tennessee, an optional protocol can be used for evaluating the lesson plan.

Analysis of teaching portfolios. Teaching portfolios are similar to but different from classroom artifacts (Goe, Bell, and Little 2008). Artifacts may be collected by the evaluator; they represent what is currently happening in the classroom. In contrast, portfolios are developed by the teacher and reflect a sample of the teacher's work over time. Using this definition, none of the five states included teaching portfolios as an element of their evaluation systems.

Self-report of teacher practice. Self-report of teacher practice allows teachers to report what is happening in their classroom (Goe, Bell, and Little 2008). Data for the self-report can be drawn from surveys, teaching records (such as teacher journals or teacher tracking of their practice and student behaviors), or interviews. All five states include a

self-report component in their teacher evaluation systems, but the formats vary. North Carolina, for example, requires teachers to rate themselves on the same rubric used in the regular evaluation and to discuss their self-evaluation at conferences with the person evaluating them throughout the year. Tennessee requires teachers to complete written self-assessments.

Student ratings of teacher performance. Students sometimes are asked to evaluate their teachers. They may be asked to assess their teacher's presentation of content, classroom management, or general approach to instruction. No states in this report included student ratings of their teachers in their evaluation systems.

Value-added models. None of the five states use value-added models (that is, student growth data) to evaluate teachers, but both Delaware and Tennessee had plans to begin doing so in 2011/12. This work has begun in Tennessee but not yet in Delaware. Student data (defined more broadly as evidence of student learning) is, however, embedded in the evaluations in some states. Georgia requires student data, such as data on the group pass rate (the percentage of a teacher's students who meet or exceed state standards) on state-mandated achievement tests, as well as other locally determined measures, as evidence of teacher performance in two domains in the evaluation rubric. In North Carolina, no student data are required, but student work is a suggested option for part of the principal's observation or teacher's self-assessment. In Tennessee, teachers provide administrators with pre- and post-assessment data on their students. Texas requires the inclusion of the school's campus rating score (an aggregate of performance data for all students in the school) in an individual teacher's ratings.

Professional development/growth plans. Goe, Bell, and Little (2008) do not include professional development plans as one of the measures in their review. This measure is included in the current overview of performance-based teacher evaluation systems because three of the five study states

(Georgia, North Carolina, and Tennessee) include some type of professional development or growth plan as a required component of the evaluation system and consider the plan in their rating of teachers.

None of the five states use value-added models (that is, student growth data) to evaluate teachers, but both Delaware and Tennessee had plans to begin doing so in 2011/12

During the first phase of evaluation in Georgia, all teachers must develop a professional growth plan, which the evaluator reviews and approves. The plan is considered in the summative evaluation of the teacher. In North Carolina, teachers must complete a professional growth plan before the initial meeting with the principal. This plan is revisited in the summative evaluation conference at the end of the school year and considered in the final rating the teacher receives. Teachers rated proficient in all domains develop an "individual growth plan" designed to improve performance in specific domains; teachers who do not receive proficient ratings in all domains are placed on a "monitored growth plan." In Tennessee, a future growth plan is included in the comprehensive assessment as evidence of the teacher's performance on the professional growth domain and is thus considered in the rating the teacher receives. In Texas, only teachers who do not meet proficiency on their evaluation must develop a plan. Texas thus does not include a professional development or growth plan as a rated or scored measure in the teacher evaluation system.

Measurement of teaching standards. States use different language, structures, and levels of detail to describe their teaching standards. In this report, *domain* refers to the most general understanding of the knowledge and skills of teaching, such as planning or the learning environment. States also identify a series of standards for the knowledge and skills teachers should possess within a particular domain. For uniformity, this report uses *standards* to refer to the knowledge and skills teachers are expected to demonstrate.

Standards were the area of focus in the analysis of the prescribed knowledge and skills that were the subject of the performance-based teacher evaluations. Across states, these standards are generally categorized into domains (content knowledge, instruction, professional responsibilities, and so forth) (table 2). For example, Tennessee classifies its 14 standards into 6 domains. Texas classifies its 50 standards into 8 domains.

To compare standards across states, the study team developed a process by which each state's standards were compared against a single set of teaching standards. This process began by reviewing a national set of model teaching standards developed by InTASC. The InTASC standards were created in 1992 as guidance for state education agencies and districts in licensing, recertifying, and evaluating teachers. These professional

TABLE 2

Teaching domains and standards in performance-based teacher evaluation systems in study states

State	Number of standards	Domains
Delaware	20	<ol style="list-style-type: none"> 1. Planning and preparation 2. Classroom environment 3. Instruction 4. Professional responsibilities 5. Student improvement
Georgia	28	<ol style="list-style-type: none"> 1. Curriculum and planning 2. Standards-based instruction 3. Assessment of student learning 4. Professionalism 5. Student achievement
North Carolina	25	<ol style="list-style-type: none"> 1. Teachers demonstrate leadership 2. Teachers establish a respectful environment for a diverse population of students 3. Teachers know the content they teach 4. Teachers facilitate learning for their students 5. Teachers reflect on their practice
Tennessee	14	<ol style="list-style-type: none"> 1. Planning 2. Teaching strategies 3. Assessment and evaluation 4. Learning environment 5. Professional growth 6. Communication
Texas	50	<ol style="list-style-type: none"> 1. Active, successful student participation in the learning process 2. Learner-centered instruction 3. Evaluation and feedback on student progress 4. Management of student discipline, instructional strategies, time, and materials 5. Professional communication 6. Professional development 7. Compliance with policies, operating procedures, and requirements 8. Improvement of academic performance of all students

Source: Authors' analysis of publicly available state documents; see appendix A for details and for more information on each system.

practice standards, revised in 2011, are designed for all teachers (Council of Chief State School Officers 2011). Ten InTASC standards are organized into four domains: the learner and learning, content knowledge, instructional practice, and professional responsibility.

The 2011 InTASC standards were used as a set of a priori codes for reviewing the state teaching standards as described in the summative teacher evaluation rubrics. Two study team members coded each state teaching standard listed in the summative evaluation, using 1 of the 10 InTASC standards (for more information about the coding process, see appendix B).

All five states' performance-based teacher evaluation system addresses all or most of the InTASC standards (table 3). The only standard not included in more than one state's evaluation rubric is Standard 1, which specifies that teachers should demonstrate an understanding of how students learn (the rubrics of Georgia and Tennessee do not reflect this standard). This standard is the least frequently used code in all states except Delaware. Delaware's teacher evaluation rubric includes neither InTASC Standard 2, specifying that teachers should demonstrate an understanding of individual learner differences, nor InTASC Standard 5, specifying that teachers should demonstrate how to engage students in critical thinking. Tennessee's teacher evaluation rubric does not include InTASC Standard 8, specifying that teachers should demonstrate use of diverse instructional strategies.

Rating categories. All five states require evaluators to provide a summative evaluation of their teachers using a predetermined rating scale. The number of rating scales, the presence of multiple ratings at the level of each teaching standard or domain, and the method for calculating a teacher's overall rating vary across states. (For more about the teaching standards in each state and links to the rubrics or scoring forms, see the state profiles in appendix A.)

Number of rating scales. Three states (Delaware, Georgia, and Tennessee) use two rating scales, one

for a preliminary set of ratings of the standards or the domains under which these standards are organized and another for the overall evaluation of teachers (table 4). Two states (North Carolina and Texas) use a single rating scale for scoring at the level of the standards or domains and for providing a final evaluation of teachers. For example, in Delaware each domain is evaluated as satisfactory or unsatisfactory, yet the final overall rating of the teacher is "effective," "ineffective," or "needs improvement." By contrast, in North Carolina, there are five possible ratings for each domain, and the overall evaluation of the teacher is based on achieving "proficient" ratings in all of the domains; there is no separate, summative rating of the teacher.

Presence of multiple ratings at the level of each teaching standard and domain. All five states provide teachers with a summative evaluation that rates their overall performance. All states also provide a performance rating for each standard or domain. Four states (all except Delaware) use multiple rating scales to rate teachers on each standard or domain (Georgia, Tennessee, and Texas use a four-rating scale, North Carolina uses a five-rating scale). Delaware rates teachers as satisfactory or unsatisfactory at the domain level; it uses these scores to determine their overall rating on a three-rating scale.

Method for calculating overall rating. In addition to providing a rating on each of the teaching standards, Delaware, Georgia, and Tennessee compute an overall rating based on the sum of the ratings on each teaching domain. In these three states, the overall rating is based on how teachers score in each domain. These states also use separate rating scales to determine the overall rating and the rating for each domain. North Carolina and Texas do not compute an overall score based on the ratings assigned to each teaching domain.

All five states require evaluators to provide a summative evaluation of their teachers using a predetermined rating scale

TABLE 3

Interstate Teacher Assessment and Support Consortium teaching standards incorporated into evaluation rubrics in study states

Teaching domain and standard	Delaware	Georgia	North Carolina	Tennessee	Texas
The learner and learning					
Standard 1. Learner development The teacher understands how learners grow and develop, recognizing that patterns of learning and development vary individually within and across cognitive, linguistic, social, emotional, and physical areas, and designs and implements developmentally appropriate and challenging learning experiences.	✓		✓		✓
Standard 2. Learner differences The teacher uses understanding of individual differences and diverse cultures and communities to ensure inclusive learning environments that enable each learner to meet high standards.		✓	✓	✓	✓
Standard 3. Learning environments The teacher works with others to create environments that support individual and collaborative learning and that encourage positive social interaction, active engagement in learning, and self-motivation.	✓	✓	✓	✓	✓
Content knowledge					
Standard 4. Content knowledge The teacher understands the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and creates learning experiences that make the discipline accessible and meaningful for learners to assure mastery of the content.	✓	✓	✓	✓	✓
Standard 5. Application of content The teacher understands how to connect concepts and use differing perspectives to engage learners in critical thinking, creativity, and collaborative problem solving related to authentic local and global issues.		✓	✓	✓	✓
Instructional practice					
Standard 6. Assessment The teacher understands and uses multiple methods of assessment to engage learners in their own growth, to monitor learner progress, and to guide the teacher's and learner's decision-making.	✓	✓	✓	✓	✓
Standard 7. Planning for instruction The teacher plans instruction that supports every student in meeting rigorous learning goals by drawing upon knowledge of content areas, curriculum, cross-disciplinary skills, and pedagogy, as well as knowledge of learners and the community context.	✓	✓	✓	✓	✓
Standard 8. Instructional strategies The teacher understands and uses a variety of instructional strategies to encourage learners to develop deep understanding of content areas and their connections and to build skills to apply knowledge in meaningful ways.	✓	✓	✓		✓

(CONTINUED)

TABLE 3 (CONTINUED)

Interstate Teacher Assessment and Support Consortium teaching standards incorporated into evaluation rubrics in study states

Teaching domain and standard	Delaware	Georgia	North Carolina	Tennessee	Texas
Professional responsibility					
Standard 9. Professional learning and ethical practice The teacher engages in ongoing professional learning and uses evidence to continually evaluate his or her practice, particularly the effects of his or her choices and actions on others (learners, families, other professionals, and the community), and adapts practice to meet the needs of each learner.	✓	✓	✓	✓	✓
Standard 10. Leadership and collaboration The teacher seeks appropriate leadership roles and opportunities to take responsibility for student learning, to collaborate with learners, families, colleagues, other school professionals, and community members to ensure learner growth and to advance the profession.	✓	✓	✓	✓	✓
Total	8	9	10	8	10

Source: Authors' analysis of data from Council of Chief State School Officers (2011) and publicly available state documents; see appendix A for details.

STUDY LIMITATIONS

There are several limitations to this review of state-level performance-based teacher evaluation systems. First, because of the emphasis on state systems, practices in place at the district or school level, including evaluation systems that may use student growth measures, are not reported.

Second, the scope of this work was limited to systems that were operational at the time the study was conducted, in spring 2011. Particularly in light of Race to the Top grant requirements, states across the country may have rolled out state-level performance-based teacher evaluation systems since then. This study does not capture these systems.

Third, Race to the Top specifies that evaluation system reforms should focus on both teachers and principals. This study looks only at teachers.

Fourth, the study depends on states' reporting. It did not draw on documents that are not publicly available, such as internal evaluations or materials related to the training of evaluators.

DIRECTIONS FOR FUTURE RESEARCH

Further analyses are warranted in several areas. Research is needed to identify the types of knowledge and skills that are commonly observed and evaluated in performance-based systems as well as knowledge and skills that may require more attention than given in current systems. Further investigation is needed into the specific nature of the measures in place, including how systems for conducting observations vary, how teaching artifacts are analyzed, and how student data are used in the overall assessment of teachers. Information beyond what is publicly available is needed about how individuals who rate teachers are prepared and trained and what structures are in place to ensure impartiality. State education agency websites do not provide information about the practical challenges, strengths, and weaknesses of teacher evaluation systems. Research that investigates the fidelity of implementation of these systems, through interviews with key constituents in these states, would elicit critical information about the successes and challenges of state-level implementation of performance-based teacher evaluation systems.

TABLE 4

Rating categories and methods for calculating overall teacher rating in study states

State	Rating scale for each standard or domain	Overall rating scale for teacher	Method for calculating overall teacher rating
Delaware	<ul style="list-style-type: none"> Satisfactory Unsatisfactory 	<ul style="list-style-type: none"> Effective Needs improvement Ineffective 	<ul style="list-style-type: none"> An “effective” rating is given to teachers who receive a satisfactory rating in at least four of five domains. A “needs improvement” rating is given to teachers who receive a satisfactory rating of three of five domains. An “ineffective” rating is given to teachers who receive a satisfactory rating on two or fewer domains.
Georgia	<ul style="list-style-type: none"> Exemplary Proficient Emerging Not evident 	<ul style="list-style-type: none"> Satisfactory Unsatisfactory 	<ul style="list-style-type: none"> Teachers are evaluated on 26 standards that fall under 5 domains. A satisfactory rating is given to teachers who score “emerging” or higher on all five domains, based on an aggregated score for the standards within each domain. An unsatisfactory rating is given to teachers whose aggregated domain score in at least one domain translates into a “not evident” rating.
North Carolina	<ul style="list-style-type: none"> Distinguished Accomplished Proficient Developing Not demonstrated 	n.a.	<ul style="list-style-type: none"> Beginning teachers must be rated “proficient” in all five domains in order to be eligible for the Standard Professional 2 License. Probationary teachers must be “proficient” in all five domains to be recommended for career (tenure) status.
Tennessee	<ul style="list-style-type: none"> Performance level C (advanced) Performance level B (proficient) Performance level A (developing) Unsatisfactory 	<ul style="list-style-type: none"> Satisfactory Unsatisfactory 	<ul style="list-style-type: none"> A “satisfactory” rating is given to first- and second-year teachers who score above level A in at least one standard per domain for domains I–IV and at or above level A in all standards in domains V and VI.a A “satisfactory” rating is given to third-year teachers who score at level B for all standards across all domains. A “satisfactory” rating is given to professional license teachers who score at level C in at least one standard in each domain, with no standard scored below level B in any domain. An “unsatisfactory” rating is given to teachers who do not meet these expectations.
Texas	<ul style="list-style-type: none"> Exceeds expectations Proficient Below expectations Unsatisfactory 	n.a.	<ul style="list-style-type: none"> A single rating of “proficient” is given to teachers who score “exceeds expectations” or “proficient” on 80 percent of the evaluation criteria for each standard in each domain and receives no ratings below “proficient.”

n.a. is not applicable. State does not compute an overall score based on the ratings assigned to each teaching domain.

a. Domain I is planning, Domain II is teaching strategies, Domain III is assessment and evaluation, Domain IV is learning environment, Domain V is professional growth, Domain VI is communication.

Source: Authors’ analysis of publicly available state documents; see appendix A for details.

Further research on performance-based teacher evaluation systems will be needed as more states, spurred in part by the Race to the Top grants, adopt performance-based systems. A scan of systems across the country might yield very different

results a year from now. This study provides useful strategies and criteria upon which to build research tracking the evolution and implementation of performance-based teacher evaluation systems.

NOTES

1. Race to the Top grant guidelines indicate that student growth data should be a “significant factor” in performance-based evaluations but do not indicate a particular percentage. (U.S. Department of Education 2009).
2. Goe, Bell, and Little (2008) specify an evaluation by a principal. This category was broadened for this study to reflect the fact that in some states the evaluation is completed by an administrator who is not the school-level principal.

APPENDIX A STATE PROFILES

TABLE A1

Profile of Delaware's Performance-Based Teacher Evaluation System

Feature	Description
Name	Delaware Performance Assessment System II (DPAS II)
Website	www.doe.k12.de.us/csa/dpasii/default.shtml
History	Implemented statewide beginning in the 2008/09 school year. All teachers are required to participate.
Goals and purpose	Contribute to professional growth, continuous improvement, and quality assurance.
Frequency of evaluation	Annual for novice teachers (teachers with an initial license), every other year for experienced teachers (teachers with a continuing or advanced license).
Professional teaching/evaluation standards	<p>Evaluation is based on five components: planning and preparation, classroom environment, instruction, professional responsibilities, and student improvement. A detailed rubric provides four indicators for each of the five components and two to five key elements for each indicator.</p> <p>DPAS II evaluates the five components through a process that involves a series of forms, conferences, and observations. The six forms used in the process are the goals form, the pre-observation form, the formative feedback form, the professional responsibilities form, the summative evaluation form, and the improvement plan form. The five conferences include a goal-setting conference, a pre-observation conference, a post-observation conference, a summative evaluation conference, and an improvement plan conference. Novice teachers are observed at least twice a year (at least one observation is announced and one is unannounced). Experienced teachers are observed at least once a year (the observation is announced). The assessment consists of a minimum 30-minute observation, a post-observation conference, and completion of a formative feedback after the conference. Teachers must submit a detailed lesson plan for the observed lesson.</p>
Measures included	Classroom observation, administrator evaluation, self-report of teacher practice
Scoring information	Administrators at the school evaluate teachers. The Department of Education provides them with evaluator training materials, including a handbook, a website, videos, and scripts for conducting conferences. The evaluator provides a score for each of the five components, each of which has equal weight in the summative score. Based on the indicators and elements, the evaluator determines if a teacher's performance is satisfactory or unsatisfactory for each component. Based on the component scores, the teacher's summative evaluation is rated "effective," "needs improvement," or "ineffective." The evaluator then determines whether the teacher requires an improvement plan for any of the five components.
Planned reforms	The Delaware Department of Education is working to establish measures for student growth, as defined in the DPAS II revised regulations. This new component is the focus of the revision process.

Source: Authors' analysis of information from www.doe.k12.de.us/csa/dpasii/student_growth/default.shtml and www.doe.k12.de.us/csa/dpasii/ti/dpasII_TeachDPASIIIGuide.pdf.

TABLE A2

Links to key resources on Delaware's Performance-Based Teacher Evaluation System

Type of information	Document
General state education agency website and other state education agency web pages	Delaware Department of Education. DPAS Homepage (n.d.) www.doe.k12.de.us/csa/dpasii/default.shtml
	Delaware Department of Education Website. Using Student Growth to Evaluate Teacher Effectiveness in Delaware. (n.d.) www.doe.k12.de.us/csa/dpasii/student_growth/default.shtml
Guides and manuals	Delaware Department of Education. DPAS II Guide for Teachers. (August 2008) www.doe.k12.de.us/csa/dpasii/ti/dpasII_TeachDPASIIGuide.pdf
Evaluation rubrics and forms	Delaware Department of Education. DPAS Rubric. (n.d.) www.doe.k12.de.us/csa/dpasii/files/rubrics/ElementsChart-July2010.pdf
Regulations and legislation	Delaware General Assembly. Delaware Regulations. Title 14 Education: 1500 Professional Standards Board. 1511 Issuance and Renewal of Continuing License. http://regulations.delaware.gov/AdminCode/title14/1500/1511.shtml#TopOfPage State of Delaware. Delaware Code. 14 DEL CODE § 1211(b). Tier Two – Continuing licensure. http://delcode.delaware.gov/title14/c012/sc02/index.shtml
Program reports	Delaware performance appraisal system second edition pilot. Year 1 report. (June 2006) www.doe.k12.de.us/csa/dpasii/pilot_eval/Year1Report.pdf Delaware performance appraisal system second edition pilot. Year 2 report. (June 2007) www.doe.k12.de.us/csa/dpasii/pilot_eval/Year2FinalReport.pdf Delaware performance appraisal system second edition. Year 1 report. (June 2008) www.doe.k12.de.us/csa/dpasii/DPAS_II_Year_2007-2008_Report.pdf Delaware performance appraisal system second edition. Year 2 report. (June 2009) www.doe.k12.de.us/csa/dpasii/files/dpassii2finalreport.pdf
Other documents	Delaware performance appraisal system second edition pilot. Year 1 report. (June 2006) www.doe.k12.de.us/csa/dpasii/pilot_eval/Year1Report.pdf Delaware performance appraisal system second edition pilot. Year 2 report. (June 2007) www.doe.k12.de.us/csa/dpasii/pilot_eval/Year2FinalReport.pdf Delaware performance appraisal system second edition. Year 1 report. (June 2008) www.doe.k12.de.us/csa/dpasii/DPAS_II_Year_2007-2008_Report.pdf Delaware performance appraisal system second edition. Year 2 report. (June 2009) www.doe.k12.de.us/csa/dpasii/files/dpassii2finalreport.pdf Delaware Department of Education. Steps for Determining Student Growth Measures for DPAS II. (n.d.) www.doe.k12.de.us/csa/dpasii/student_growth/files/Steps_Determin_Stu_Grth_Meas.pdf

Note: All resources last retrieved July 9, 2011.

TABLE A3

Profile of Georgia's Performance-Based Teacher Evaluation System

Feature	Description
Name	Classroom Analysis of State Standards (CLASS Keys) Georgia Teacher Evaluation System
Website	www.gadoe.org/tss_teacher.aspx
History and legislation	<p>CLASS Keys replaced the Georgia Teacher Evaluation Program (GTEP) in 2010. Administrators and teachers served as co-developers by providing feedback that was used to refine the performance appraisal process.</p> <p>Georgia law requires that teachers be evaluated in part on the academic achievement of their students. The law, which was updated in 2006, allows for various ways of measuring student achievement, including the use of student data.</p>
Goals and purpose	Support teachers' work in standards-based classrooms using the Georgia Performance Standards to improve student learning, improve teacher performance, and increase accountability. This feature is both summative and formative.
Frequency of evaluation	Annual
Professional teaching/evaluation standards	<p>The teacher evaluation system is designed around five strands, or "keys," that are aligned with the School Keys (Georgia's standards for a comprehensive system of school improvement and support): curriculum and planning, standards-based instruction, assessment of student learning, professionalism, and student achievement. CLASS Keys has a Crosswalk that lists standards and elements for the five strands. The strands are broken into performance standards and elements with rubrics that have accompanying evidence and artifacts.</p> <p>CLASS Keys includes three phases. In Phase 1, teachers self-assess their performance using the continuum of improvement rubrics. After reflecting on their areas of strength and areas for growth, they develop a draft of their professional growth plan. At the pre-evaluation conference, the evaluator reviews and approves this plan. Student achievement targets are set, and expectations are clarified regarding elements, duties, and responsibilities.</p> <p>In Phase 2, evaluators collect evidence by conducting short, unannounced classroom observations to assess a few of the elements. Later, evaluators conduct a longer, announced classroom observation to assess as many elements as possible. Evaluators also collect evidence from other sources, including conferences, meetings, planning and professional learning sessions, and student and teacher products.</p> <p>In Phase 3, evaluators score the teacher's annual performance on each element by reviewing all of the evidence collected during the year, using the continuum of improvement rubrics.</p>
Measures included	Classroom observation, administrator evaluation, analysis of classroom artifacts, self-report of teacher practice, professional development/growth plans
Use of student data	Student data are collected and assessed for two of the five strands (assessment of student learning and student achievement). These data may include student data records, data on the group pass rate (the percentage of a teacher's students who met or exceeded state standards) on state-mandated academic achievement tests, and state- and district-level student data on the percentage of students who met or exceeded state standards on state-mandated achievement measures.
Scoring information	Online training modules, intended to inform and train evaluators, are provided for each section of the evaluation process. For the student achievement elements and strand, scoring is based on student achievement gains by a teacher's students compared with goals set earlier in the year in phase 1. The teacher's performance on the Georgia Teacher Duties and Responsibilities (GTDR) is also reviewed. If overall performance on the GTDR is "satisfactory" and all five strands are scored at the "emerging" level or higher, the teacher receives an overall score of "satisfactory" for the annual evaluation. If the teacher's overall performance is "unsatisfactory," a Professional Development Plan is required.
Planned reforms	There is no information on changes to the current system.

Source: Authors' analysis of information from www.doe.k12.ga.us/DMGetDocument.aspx/CK%20Process%20Guide%2010-6-10.pdf?p=6CC6799F8C1371F6EB4CBCF928752B238AF6CAF7B70DA773616C26BABA1D9AFC&Type=D, www.doe.k12.ga.us/tss_teacher.aspx?PageReq=TSSTrainingModules, www.doe.k12.ga.us/DMGetDocument.aspx/CK%20Standards%2010-29-08.pdf?p=6CC6799F8C1371F60C8684DFDC96C1C9E173A927D7D04E1B1E862FC762CCF7F9&Type=D, <http://law.onecle.com/georgia/20/20-2-210.html>, www.gadoe.org/DMGetDocument.aspx/CK%20Crosswalk%204-7-2011.pdf?p=6CC6799F8C1371F616F0C83A91176799CD3986AAF3BF5EE7EEB1D7BA163B7D5D&Type=D, and www.doe.k12.ga.us/DMGetDocument.aspx/CK%20Standards%2010-29-08.pdf?p=6CC6799F8C1371F60C8684DFDC96C1C9E173A927D7D04E1B1E862FC762CCF7F9&Type=D.

TABLE A4

Links to key resources on Georgia's Performance-Based Teacher Evaluation System

Type of information	Document
General state education agency website and other state education agency web pages	Georgia Department of Education. Teacher and Leader Quality home page. (2010) www.gadoe.org/tss_teacher.aspx
	Georgia Department of Education. CLASS Keys Training Modules. (2010) www.doe.k12.ga.us/tss_teacher.aspx?PageReq=TSSTrainingModules
Guides and manuals	Georgia Department of Education. CLASS Keys Overview and Guide. (April 2009) www.doe.k12.ga.us/DMGetDocument.aspx/CK%20Standards%2010-29-08.pdf?p=6CC6799F8C1371F60C8684DFDC96C1C9E173A927D7D04E1B1E862FC762CCF7F9&Type=D
	Georgia Department of Education. CLASS Keys Process Guide. (July 2010) www.doe.k12.ga.us/DMGetDocument.aspx/CK%20Process%20Guide%2010-6-10.pdf?p=6CC6799F8C1371F6EB4CBCF928752B238AF6CAF7B70DA773616C26BABA1D9AFC&Type=D
	Georgia Department of Education. School Keys Manual. (May 2007) www.doe.k12.ga.us/DMGetDocument.aspx/SCHOOL%20KEYS%20FINAL%205-29-07.pdf?p=6CC6799F8C1371F6175E5B6E474BB7C617F852E1ADE57E7942B6D677375DA861&Type=D
Evaluation rubrics and forms	Georgia Department of Education. CLASS Keys Forms. (2010) www.doe.k12.ga.us/tss_teacher.aspx?PageReq=TSSClassKeysForms
Regulations and legislation	Georgia Code - Education - Title 20, Section 20-2-210 (2006) http://law.onecle.com/georgia/20/20-2-210.html
Other documents	Georgia Department of Education. CLASS Keys Crosswalk. (April 2011) www.gadoe.org/DMGetDocument.aspx/CK%20Crosswalk%204-7-2011.pdf?p=6CC6799F8C1371F616F0C83A91176799CD3986AAF3BF5EE7EEB1D7BA163B7D5D&Type=D

Note: All resources last retrieved July 9, 2011.

TABLE A5

Profile of North Carolina's Performance-Based Teacher Evaluation System

Feature	Description
Name	North Carolina Teacher Evaluation Process (NCTEP)
Website	www.ncpublicschools.org/profdev/training/teacher/
History	Phase I was piloted in August 2008. Phase II, implemented in August 2009, added more districts. Phase III, implemented in August 2010, added all remaining districts in the state, making NCTEP a statewide system in the 2010/11 school year.
Goals and purpose	Assess performance on standards, and provide a tool for development of individual practitioner growth.
Frequency of evaluation	All teachers complete an orientation, a self-assessment, and a professional development plan annually. Career status (tenured) teachers are required to participate in a formal evaluation cycle, including observations, at least once every five years. (In the off-cycle years, local education agencies may require components of these evaluations.) For probationary (untenured) teachers, the formal evaluation process, including observations and a summary evaluation conference, is annual.
Professional teaching/evaluation standards	<p>North Carolina uses five standards, adopted in 2007: teachers demonstrate leadership, teachers establish a respectful environment for a diverse population of students, teachers know the content they teach, teachers facilitate learning for their students, and teachers reflect on their practice. Each standard includes several indicators (a total of 25 across the 5 standards) that delineate what a teacher should know and be able to do.</p> <p>NCTEP includes eight components: training, orientation, self-assessment, pre-observation conference, observation, post-observation conference, summary evaluation conference, and professional development plan. Before the first observation, the principal meets with the teacher to discuss the teacher's self-assessment, the teacher's most recent professional growth plan, and the lesson or lessons to be observed. Pre-observation conferences are required only for the first observation. A formal observation lasts at least 45 minutes or an entire class period. The principal conducts at least three observations of probationary teachers. Peer evaluators conduct one formal evaluation of probationary teachers. During the year in which a career status teacher participates in a summative evaluation, the principal conducts at least three observations, including at least one formal observation. No later than 10 school days after each formal observation, the principal conducts a post-observation conference in which the principal and teacher discuss and document the strengths and weaknesses of the teacher's performance. Before the end of the school year, the principal conducts a summary evaluation conference with the teacher, in which the principal and teacher discuss the teacher's self-assessment, the teacher's most recent professional growth plan, the components of the NCTEP completed during the year, classroom observations, artifacts submitted or collected during the evaluation process, and other evidence of the teacher's performance according to the rubric.</p> <p>Teachers rated at least "proficient" on all standards on the teacher summary rating form develop an individual growth plan designed to improve their performance on specific standards and elements. Teachers are placed on monitored growth plans if they are rated "developing" on one or more standards on the teacher summary rating form and are not recommended for dismissal, demotion, or nonrenewal.</p>
Measures included	Classroom observation, administrator evaluation, self-report of teacher practice, professional development/growth plans
Use of student data	No student data are required; however, teachers can provide student work as part of their self-assessment and assessment by the evaluator.
Scoring information	<p>The principal or a designee conducts the evaluation. Observations are conducted by observers who are licensed administrators or who have a supervisory certificate. Anyone trained on the evaluation process may be a peer observer.</p> <p>Teachers receive ratings of "developing," "proficient," "accomplished," "distinguished," or "not demonstrated." Third-year teachers must have an overall rating of "proficient" or higher on each of the five standards in order to be recommended for a Standard Professional License II and continued employment. Probationary teachers in their fourth year of teaching or after one year of probationary status must receive an overall rating of "proficient" or higher on the five standards to be recommended for career (tenured) status.</p>

(CONTINUED)

TABLE A5 (CONTINUED)

Profile of North Carolina's Performance-Based Teacher Evaluation System

Feature	Description
Planned reforms	There is no information on changes to the current system.

Source: Authors' analysis of information from www.ncpublicschools.org/docs/profdev/training/teacher/teacher-evaluation.pdf, www.ncpublicschools.org/docs/profdev/training/teacher/important-points.pdf, www.ncga.state.nc.us/EnactedLegislation/Statutes/HTML/BySection/Chapter_115C/GS_115C-335.html, www.ncpublicschools.org/docs/profdev/training/teacher/lea-eval-process.pdf, and www.ncpublicschools.org/profdev/training/teacher/.

TABLE A6

Links to key resources on North Carolina's Performance-Based Teacher Evaluation System

Type of information	Document
General state education agency website and other state education agency web pages	North Carolina Department of Public Instruction. Professional Development. (2009) www.ncpublicschools.org/profdev/training/teacher/
Guides and manuals	North Carolina Teacher Evaluation Process Manual. (2009) www.ncpublicschools.org/docs/profdev/training/teacher/teacher-eval.pdf North Carolina State Board of Education Policy Manual. (2008) www.ncpublicschools.org/docs/profdev/training/teacher/teacher-evaluation.pdf
Evaluation rubrics and forms	Details of NC Teacher Evaluation Instrument. (n.d.) www.ncpublicschools.org/docs/profdev/training/teacher/lea-eval-process.pdf Teacher Evaluation Rubric. (n.d.) www.ncpublicschools.org/docs/profdev/training/teacher/materials/eval-rubric.doc
Regulations and legislation	North Carolina Statute. 115C-335. (1998) www.ncga.state.nc.us/EnactedLegislation/Statutes/HTML/BySection/Chapter_115C/GS_115C-335.html
Other documents	North Carolina Professional Educator Evaluation Systems. Training Materials. (n.d.) www.ncpublicschools.org/profdev/training/ North Carolina Professional Teaching Standards. (2006) www.ncpublicschools.org/docs/profdev/standards/teachingstandards.pdf

Note: Resources last retrieved July 9, 2011.

TABLE A7

Profile of Tennessee’s Performance-Based Teacher Evaluation System

Feature	Description
Name	Framework for Evaluation and Professional Growth
Website	http://state.tn.us/education/frameval/index.shtml#files
History	The Tennessee State Board of Education approved a teacher evaluation process in 1997 that became effective statewide in July 2000. In 2004, the board approved revisions to the original model, with the goal of improving rigor and structuring the system to align with the No Child Left Behind Act of 2001. The statewide framework was revised again in 2009.
Goals and purpose	Encourage teachers to move beyond their current level of performance by focusing on student growth and self-reflection on areas for their own growth and school improvement.
Frequency of evaluation	Tenured teachers (teachers with a professional license) are evaluated every five years, with a minimum of two observations. Novice teachers (teachers with an apprentice license) are evaluated annually, with a minimum of three observations during the first two years and two observations in the third year.
Professional teaching/evaluation standards	<p>Six domains (planning, teaching strategies, assessment and evaluation, learning environment, professional growth, and communication) cover 14 indicators of teacher behaviors and characteristics. Each indicator includes criteria that are directly aligned with four rating levels (“developing,” “proficient,” “advanced,” and “unsatisfactory”). Evaluators must be trained on the Framework for Evaluation model. Training consists of three days instruction delivered over several months.</p> <p>The Framework for Evaluation and Professional Growth includes a self-assessment, discussion of previously collected information, an observation process, a planning information record, classroom notes, a reflecting information record, an appraisal record, and an educator information record. The inclusion of a unit plan/lesson plan is optional, and the requirement may vary locally. A summative process includes analysis of data, identification of performance levels, sharing of evaluation data, and preparation of a future growth plan.</p>
Measures included	Classroom observation, administrator evaluation, self-report of teacher practice, analysis of classroom artifacts, professional development/growth plans.
Use of student data	No student growth data are required, although teachers are required to use student achievement data as a means of communicating progress to students and informing their instructional practice.
Scoring information	Teachers are rated on each of the 14 indicators. There are four rating categories: unsatisfactory; performance level A (developing); performance level B (proficient); and performance level C (advanced). Expectations depend on a teacher’s level of experience. First- and second-year teachers are expected to score above Level A on at least one indicator for domains I–IV, with all indicators at level A in Domains V and VI (The domains are planning, teaching strategies, assessment and evaluation, learning environment, professional growth, and communication). Third-year teachers are expected to score at level B for all indicators across all domains. Professional license teachers are expected to score at level C in at least one indicator in each domain, with no indicators scored below level B in any domains.
Planned reforms	On January 25, 2010, the governor of Tennessee signed the Tennessee First to the Top Act. The act requires annual evaluation of teachers and the use of value-added model scores in the evaluation process. Beginning July 1, 2011, all teachers in Tennessee are evaluated annually, with four observations by the principal (two per semester), at least half of them unannounced. A pre-conference is required for all announced observations; a post-conference is required for all observations.

Source: Authors’ analysis of information from <http://state.tn.us/sos/rules/0520/0520-02/0520-02-01.pdf>, <http://state.tn.us/education/frameval/index.shtml#files>, http://state.tn.us/education/frameval/doc/comprehensive_assessment.pdf, <http://state.tn.us/sbe/2010January13pdfs/Minutes%20of%20January%2013%202010%20Special%20Session.pdf>, <http://state.tn.us/sbe/2010January13pdfs/Minutes%20of%20January%2013%202010%20Special%20Session.pdf>, and www.tn.gov/firsttothetop/FieldTest1pager.pdf.

TABLE A8

Links to key resources on Tennessee's Performance-Based Teacher Evaluation System

Type of information	Document
General state education agency website and other state education agency web pages	Tennessee Department of Education. Framework for Evaluation and Professional Growth. (n.d.) http://state.tn.us/education/frameval/index.shtml#files
	Tennessee Department of Education. First to the Top. Teacher and Principal Evaluation. (n.d.) www.tn.gov/firsttothetop/programs-committee.html
Guides and manuals	Framework for Evaluation and Professional Growth. Comprehensive Assessment. (2009) http://state.tn.us/education/frameval/doc/comprehensive_assessment.pdf
Regulations and legislation	Rules of the State Board of Education. Chapter 0520-2-1 Evaluations. (January 2008) http://state.tn.us/sos/rules/0520/0520-02/0520-02-01.pdf
Other documents	State Board of Education Meeting Minutes. (January 13 2010) http://state.tn.us/sbe/2010January13pdfs/Minutes%20of%20January%2013%202010%20Special%20Session.pdf

Note: Resources last retrieved July 9, 2011.

TABLE A9

Profile of Texas's Performance-Based Teacher Evaluation System

Feature	Description
Name	Professional Development and Appraisal System (PDAS)
Website	http://www5.esc13.net/pdas/
History and legislation	Senate Bill 1, passed in 1995, required the commissioner of education to develop a recommended appraisal system for Texas teachers, with input from teachers and other professionals. Since the 1997/98 school year, all school districts have had two choices in selecting a method by which to appraise teachers: a teacher-appraisal system recommended by the commissioner of education or a local teacher-appraisal system. The commissioner's recommended teacher-appraisal system, the PDAS, was developed in accordance with TEC §21.351. Provisions were adopted or amended August 1, 1997; April 15, 1999; July 31, 2001; May 31, 2004; and February 17, 2010. The superintendent of each school district, with the approval of the school district's board of trustees, may select the PDAS. A school or school district that prefers to select or develop an alternative teacher-appraisal system must follow TEC §21.352.
Goals and purpose	Enhance student learning through the professional development of educators.
Frequency of evaluation	At least once each school year. Since 2003, legislation allows districts to adopt policies at the local level that modify the annual appraisal schedule for qualifying teachers, as long as an appraisal is performed at least once every five years. A teacher must be rated as at least proficient on each PDAS domain to be eligible for less frequent appraisals. A school district may choose to review the written agreement with a teacher annually. However, at the end of the school year, the district may modify appraisal options through board policy and make changes to expectations for appraisals that apply to all teachers, regardless of a teacher's participation in the appraisal option the previous year.

(CONTINUED)

TABLE A9 (CONTINUED)

Profile of Texas's Performance-Based Teacher Evaluation System

Feature	Description
Professional teaching/evaluation standards	<p>Evaluation covers eight domains reflecting the teacher proficiencies described in Learner-Centered Schools for Texas: A Vision of Texas Educators, approved in 1994 by the State Board for Educator Certification. The domains are active, successful student participation in the learning process; learner-centered instruction; evaluation and feedback on student progress; management of student discipline, instructional strategies, time, and materials; professional communication; professional development; compliance with policies, operating procedures, and requirements; and improvement of academic performance of all students in the school.</p> <p>Evaluation is based on classroom observation and walkthroughs. The system requires a minimum of one observation of at least 45 minutes, plus additional observations and walkthroughs as necessary. Teachers whose performance is appraised as less than proficient in any domain must be given the opportunity to improve their performance through the development of an intervention plan.</p> <p>The PDAS also provides for teachers' input into their own appraisal ratings, especially in domain VI (professional development) and domain VIII (efforts to improve academic performance), through the inclusion of the teacher self-report form, which teachers can use to submit concrete examples of their best work for consideration in the appraisal process.</p>
Measures included	Classroom observation, administrator evaluation, self-report of teacher practice, professional development/growth plans
Use of student data	<p>Student performance must be included in each teacher's appraisal. Domain VIII (improvement of academic performance of all students on the campus) addresses the student performance link. The campus (school) performance rating, which incorporates the current state accountability system report and adequate yearly progress indicators, is use to score criterion 10 of domain VIII. The other nine criteria in domain VIII relate to a teacher's focus on various school goals associated with the improvement of academic performance for all students.</p>
Scoring information	<p>The evaluator (called an appraiser) is usually the teacher's supervisor, although principals, assistant principals, other administrators designated by supervisory staff, or other professionals hired by the superintendent can also serve as appraisers. Beginning administrators seeking certification as PDAS appraisers must complete a 36-hour training.</p> <p>Teachers are given one of four ratings ("exceeds expectations," "proficient," "below expectations," "unsatisfactory") in each of the eight domains. Each domain is scored independently; there are no aggregate scores.</p> <p>Appraisers first identify evidence related to the 51 critical attributes of the criteria, as specified in the PDAS Appraisal Framework and the Observation Summary. They then view the evidence in light of both quality and quantity. Quality focuses on the "strength, impact, variety, and alignment" (SIVA) of the teaching behavior and how it relates to student success. Quantity relates to the frequency and number of students for which the teaching behavior resulted in student learning. Appraisers use the PDAS Appraisal Framework, Scoring Framework, THE Performance Level Standards (SIVA), and the Scoring Criteria Guide to make performance-level decisions.</p>
Planned reforms	There is no information on changes to the current system.

Source: Authors' analysis of information from www.hempstead.isd.esc4.net/hisd-web/Information-Resources/InfoAttachments/PDASTeacherManual2005.pdf, www5.esc13.net/pdas/docs/forms/ScriptingForm.pdf, www5.esc13.net/pdas/docs/forms/tsrf.pdf, <http://ritter.tea.state.tx.us/rules/tac/ch150aa.html>, and <http://www5.esc13.net/pdas/docs/LearnerCenteredSchools.pdf>.

TABLE A10

Links to key resources on the Texas's Professional Development and Appraisal System

Type of information	Document
General state education agency website and other state education agency web pages	Texas Education Agency. Professional Development and Appraisal System. (n.d.) http://www5.esc13.net/pdas/
Guides and manuals	Professional Development and Appraisal System. Teacher Manual. (2005) www.hempstead.isd.esc4.net/hisd-web/Information-Resources/InfoAttachments/PDASTeacherManual2005.pdf
Scoring guides	2004 PDAS Revision. Scoring Criteria Guide. (2004) http://www5.esc13.net/pdas/docs/forms/ScoringCriteriaGuide.pdf Professional Development and Appraisal System. Scoring Factors and Performance Level Standards. (2004) http://www5.esc13.net/pdas/docs/forms/SIVACChart.pdf
Evaluation rubrics and forms	Development and Appraisal System. Observation/Scripting/Documentation Form. (2004) http://www5.esc13.net/pdas/docs/forms/ScriptingForm.pdf Professional Development and Support Teacher Self-Report Form. (2001) http://www5.esc13.net/pdas/docs/forms/tsrf.pdf 2004 PDAS Revision Appraisal Framework. (2004) http://www5.esc13.net/pdas/docs/forms/Framework.pdf 2004 Revision. Observation Summary. (2004) http://www5.esc13.net/pdas/docs/forms/ObservationSummary.pdf
Regulations and legislation	Texas. Chapter 150. Commissioner's Rules Concerning Educator Appraisal. (n.d.) http://ritter.tea.state.tx.us/rules/tac/ch150aa.html
Other documents	Learner-Centered Schools for Texas: A Vision of Texas Educators. (1997) http://www5.esc13.net/pdas/docs/LearnerCenteredSchools.pdf Frequently Asked Questions for Professional Development and Appraisal System (PDAS). (n.d.) http://www5.esc13.net/pdas/docs/PDASFAQ.pdf

Note: All resources last retrieved July 9, 2011.

**APPENDIX B
METHODOLOGY**

This appendix describes the data search criteria and methodology used in this study.

Identifying states for inclusion

To develop the inclusion criteria for this study, the study team began with the definition of performance-based teacher evaluation systems used in the Race to the Top federal grant guidelines. They augmented these guidelines based on scholarly consensus regarding multiple measures. After reviewing state department of education websites for all 50 states, the study team revised the criteria, because no state used student growth data and only one state required annual reviews of teachers (two elements of the Race to the Top grant guidelines).

The criteria used to identify states for inclusion in this study included the following:

- The evaluation system was required for practicing general educators.
- The system was operational statewide as of the 2010/11 school year.
- The system included multiple rating categories.
- The system used multiple measures.

TABLE B1
Terms used to search for state-level performance-based teacher evaluation systems

Broad evaluation terms	Performance-related evaluation terms
Educator effectiveness	Performance assessment
Educator evaluation	Performance-based
State teacher evaluation	Performance evaluation
Teacher effectiveness	Portfolio
Teacher evaluation	Student growth
Teacher evaluation system	Value-added

Source: Authors' selection based on the criteria used to identify states for inclusion in the study.

Based on these criteria, the study team created a list of terms that guided the search of publicly available documents on performance-based teacher evaluation systems (table B1).

Each team member conducted a search of about 10 states, searching state education agency websites and using Google searches of the search terms with the state's name included. Study team members used the information found on state websites, including downloadable documents such as evaluation rubrics and manuals, to complete table B2. On a shared Google document, they noted what they found from scanning state websites and documents, included links to pertinent information, raised questions for further investigation, and noted how the information on each state matched up to the inclusion criteria. A different team member then reviewed

TABLE B2
Form used to record information on states' performance-based teacher evaluation systems

State	Reviewer 1	Reviewer 2	Required statewide	Required for practicing general educators	Uses multiple measures	Includes student growth data	Includes multiple rating categories	Requires annual evaluation for novice teachers	Requires evaluation for experienced teachers
Alabama									
Alaska									
...									
Wyoming									

Source: Authors' analysis of publicly available state documents.

the first team member's comments and conducted a similar search to respond to questions and confirm the accuracy of the information. The scan identified 38 states that did not have evaluation systems in place for all educators; 7 states that had evaluation systems in place but did not meet other criteria; and 5 states (Delaware, Georgia, North Carolina, Tennessee, and Texas) that met all study criteria (table B3).

Collecting data

Members of the study team reviewed the state education agency websites of all states that met the

study criteria, examining the following types of documents:

- General state education agency website and other state education agency web pages.
- Evaluation guides and manuals.
- Evaluation rubrics and forms.
- Program reports.
- Other documents and additional information, such as frequently asked questions and presentations about evaluations, including training materials for teachers, principals, and evaluators.

TABLE B3

Overview of performance-based teacher evaluation systems in study states

State	System	Start date	Required statewide	Required for practicing general educators	Uses multiple measures	Includes student growth data	Includes multiple rating categories	Requires annual evaluation for novice teachers	Requires evaluation for experienced teachers
Delaware	Delaware Performance Assessment System II	2008/09	✓	✓	✓		✓	✓	Every other year
Georgia	Classroom Analysis of State Standards Keys Teacher Evaluation System	2010/11	✓	✓	✓		✓	✓	✓
North Carolina	North Carolina Teacher Evaluation Process	2010/11	✓ ^a	✓	✓		✓	✓	Partial evaluation annually
Tennessee	Framework for Evaluation and Professional Growth	2000 (revised in 2004 and 2009)	✓	✓	✓		✓	✓	Every five years
Texas	Professional Development Appraisal System	1997/98 (revised in 1999, 2001, 2004, and 2010)	✓ ^a	✓	✓		✓	At discretion of district ^b	At discretion of district

a. In North Carolina and Texas, state regulations permit districts to create their own evaluation system as long as it is comparable to the system the state has developed and meets all regulatory requirements.

b. Although districts are free to develop their own timeline for the evaluations, all teachers must be rated at least proficient on all domains to be eligible for appraisals that do not occur annually.

Source: Authors' analysis of publicly available state documents.

To answer the first research question, the study team collected the following information from each state's website:

- Evaluation system name.
- Website.
- History and legislation.
- Goals and purpose.
- Frequency of evaluation.
- Professional teaching standards/evaluation standards.
- Process and evaluation measures.
- Use of student data.
- Scoring information (multiple rating categories, rubrics, and so forth).
- Planned system changes.
- Links to key resources.

To answer the second research question, the study team collected information on measures states use to evaluate teachers, the teaching standards the evaluations were designed to measure, and the rating categories used to evaluate teachers.

In an effort to obtain comparable information, such as scoring guides, information about evaluator training, and evaluations of each system, the study team contacted certification and licensure, instructional leadership, and other departments in each state by email (box B1). Each email also included questions specific to each state, which were included in an effort to clarify questions that remained after initial information had been gathered. No additional information was collected from these contacts.

Analyzing the data

For each state, a member of the study team created a profile of the state system based on the information and documents available on its website. To ensure the relevance and accuracy of all information, a second team member checked all links and read all documents cited in the profile.

The study team examined the similarities and differences in evaluation systems in terms of the measures states used to evaluate teachers, the teaching standards the evaluations were designed

BOX B1

Text of email sent to state officials requesting information statewide performance-based teacher evaluation systems

Dear _____,

I am a researcher for the Regional Educational Laboratory-Northeast and Islands (www.relnei.org/home.php), based in Newton, MA. The REL-NEI is one of a network of ten laboratories that provides educators and policymakers with access to high-quality scientifically valid education research through applied research and development projects, randomized controlled trial studies, dissemination of research findings, and related technical assistance activities. The REL Program is funded by the Institute of Education Sciences at the U.S. Department of Education.

The REL-NEI is currently conducting a review of existing statewide performance-based teacher evaluation systems across the country. We have identified STATE as one of the states for our review, based on the information available on the state website. Could you confirm that all of the information about your teacher evaluation system is up-to-date on your website? If not, could you provide additional information pertaining to teacher evaluation that may not yet be available on your website?

to measure, and the categories used to rate teachers. Each of these subquestions required a different data analysis strategy.

Evaluation measures. The study team created a data collection matrix based on Goe, Bell, and Little's (2008) categorization of evaluation measures. These measures include classroom observations, evaluations by principals, analysis of classroom artifacts, analysis of teaching portfolios, teacher self-reports of practice, student ratings of teacher performance, and value-added strategies. In some cases, states used tools that did not correspond exactly with the categories identified by Goe, Bell, and Little (2008); in these cases, the study team modified some of the categories. Two study team members reviewed the table and the original documents to check for

accuracy. Once the table was completed, the study team identified similarities and differences across states for each of category.

Teaching standards. For each state, the study team compiled a list of teaching standards and domains from the summative teaching evaluation rubrics or scoring forms collected. It used a list of a priori codes derived from the 2011 Interstate Teacher Assessment and Support Consortium (InTASC) Teaching Standards to compare standards across states (see table 4). To code each of the state's standards using the InTASC codes, the study team created a data analysis matrix listing the state's standards on the left and one of the 10 InTASC codes on the right.

The coding process was iterative. After coding a state, study team members discussed the use of the codes and established an additional coding rule in which they addressed cross-cutting themes identified by InTASC. InTASC identified these themes (such as communication and technology) as key concepts that were reflected in more than one standard but not explicitly articulated as a standard. A chart in the InTASC document provided all of the standards associated with each cross-cutting theme. In coding state standards documents, the study team found that some cross-cutting themes were articulated as state teaching standards. In these cases, it determined which InTASC standards were most relevant to each cross-cutting theme. For example, in Tennessee, one teaching standard requires teachers to communicate clearly and correctly with students, parents, and other stakeholders. Although no InTASC standard maps directly to this requirement, InTASC Standard 3, regarding learning environments, relates most closely to it. This standard in Tennessee was thus coded as InTASC Standard 3.

Two team members coded all five states, using the following process. First, each team member independently evaluated each state, based on each state's summative teacher evaluation rubric or scoring form, coding each standard with a single InTASC standard. For each standard, InTASC provides substantial detail about the knowledge, disposition, and performance expected of teachers. In the coding process,

the study team relied not only on the specificity of each standard but also on this additional detail. The document that includes this information is available on the Council of Chief State School Officers website, at www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf.

Second, the study team measured interrater reliability using Cohen's Kappa statistic, a measure that is more robust than simple percent agreement because it accounts for chance agreement (Cohen 1968). The purpose of estimating the Kappa statistic was to get a sense of the suitability of using the InTASC standards as a coding scheme. The following formula was used to calculate the Kappa statistic:

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

where $Pr(a)$ is the relative observed agreement among raters and $Pr(e)$ is the hypothetical probability of chance agreement. If the raters are in complete agreement, $K = 1$. Interrater reliability ranged from .67 to .78 (table B4).

Third, the two team members met and discussed their coding. When they encountered disagreement in their choices, they discussed the issue and went back to the InTASC code definitions for guidance. The two team members could not reach consensus regarding a single InTASC code to use for four items (one item in Texas, one in Tennessee, and two in Georgia).

Fourth, a third team member resolved the disagreement on the four unreconciled items.

TABLE B4

Interrater reliability of coding of state performance-based teacher evaluation systems

State	Kappa statistic
Delaware	.71
Georgia	.71
North Carolina	.78
Tennessee	.68
Texas	.67

Source: Authors' analysis using Cohen's Kappa statistic.

REFERENCES

- Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.
- Braun, H. (2005). *Using student progress to evaluate teachers: a primer on value added models*. Princeton, NJ: ETS: Retrieved March 18, 2011, from www.ets.org/Media/Research/pdf/PICVAM.pdf.
- Coggshall, J., Max, J., and Bassett, K. (2008). *Key issue: using performance-based assessment to identify and support high-quality teachers*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved July 5, 2011, from www.wested.org/schoolturnaroundcenter/docs/coggshall-assessment.pdf.
- Cohen, J. (1968). Weighed kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70 (4): 213–220.
- Council of Chief State School Officers (2011). *Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards*. Washington, DC: CCSSO: Retrieved June 6, 2011, from www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf.
- Daley, G., and Kim, L. (2010). *A teacher evaluation system that works: a working paper*. Santa Monica, CA: National Institute for Excellence in Teaching. Retrieved June 27, 2011, from www.tapsystem.org/publications/wp_eval.pdf.
- Goe, L., Bell, C., and Little, O. (2008). *Approaches to evaluating teacher effectiveness: a research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved November 17, 2010, from www.tqsource.org/link.php.
- Gordon, R., Kane, T. J., and Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: The Brookings Institution. Retrieved November 17, 2010, from www.brookings.edu/~media/Files/rc/papers/2006/04education_gordon/200604hamilton_1.pdf.
- Guyatt, G., and Rennie, D. (Eds.). (2002). *Users' guide to the medical literature: a manual for evidence based clinical practice*. Chicago, IL: American Medical Association.
- Hanushek, E. and Rivkin, S. (2010). *Using value-added measures of teacher quality*. Washington, DC: National Center for Analysis of Longitudinal Data in Educational Research; The Urban Institute. Retrieved December 5, 2010, from www.urban.org/UploadedPDF/1001371-teacher-quality.pdf.
- Harris, D. (2008). *Value-added measures of education performance: clearing away the smoke and mirrors, Policy Brief 10-4*. Stanford, CA: Stanford University; PACE. Retrieved November 11, 2010, from www.stanford.edu/group/pace/PUBLICATIONS/PB/PACE_BRIEF_OCT_2010.pdf.
- Heneman, H. G., III, Milanowski, A., Kimball, S. M., and Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay. Policy brief RB-45*. Philadelphia, PA: Consortium for Policy Research in Education. Retrieved November 17, 2010, from <http://cpre.wceruw.org/publications/rb45.pdf>.
- Jacobs, B. A., and Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101–136.
- Kane, T., Rockoff, J., and Staiger, D. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City*. NBER Working Paper 12155. Cambridge, MA: National Bureau of Economic Research. Retrieved November 15, 2010, from www.gse.harvard.edu/news/features/kane/nycfellowsmarch2006.pdf.
- Learning Point Associates (2010). *Evaluating teacher effectiveness: emerging trends reflected in the State Phase 1 Race to the Top applications*. Naperville, IL: Learning Point Associates. Retrieved April 10, 2011, from www.learningpt.org/pdfs/RttT_Teacher_Evaluation.pdf.

- Measures of Effective Teaching Project (2010). *Learning about teaching: initial findings from the measures of effective teaching project*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved July 5, 2011, from <http://s3.documentcloud.org/documents/18327/met-research-paper.pdf>.
- Milanowski, M. T., Heneman, H. G. III., and Kimball, S. T. (2009). *Review of teaching performance assessments for use in human capital management*. Madison, WI: Consortium for Policy Research in Education, Wisconsin Center for Education Research, University of Wisconsin, Madison. Retrieved July 5, 2011, from www.eric.ed.gov/PDFS/ED506953.pdf.
- National Council on Teacher Quality (2009). *National council on teacher quality state teacher policy yearbook 2009*. Washington, DC: National Council on Teaching Quality. Retrieved November 17, 2010, from www.nctq.org/stpy09/.
- Toch, T., and Rothman, R. (2008). *Rush to judgment: teacher evaluation in public education*. *Education Sector Reports*. Washington, DC: Education Sector, Retrieved March 10, 2011, from www.educationsector.org/sites/default/files/publications/RushToJudgment_ES_Jan08.pdf.
- U.S. Department of Education (2009). *Race to the Top: executive summary*. Washington, DC: U.S. Department of Education. Retrieved November 17, 2010, from <http://www2.ed.gov/programs/racetothetop/index.html>.
- Weisberg, D., Sexton, S., Mulhern, J., and Keeling, D. (2009). *The widget effect: our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teachers Project. Retrieved July 5, 2011, from <http://widgeffect.org/downloads/TheWidgetEffect.pdf>.