



Using Multiple Evaluation Measures to Improve Teacher Effectiveness

State Strategies from Round 2 of No Child Left Behind Act Waivers

Glenda L. Partee December 2012



Using Multiple Evaluation Measures to Improve Teacher Effectiveness

State Strategies from Round 2 of No Child Left Behind Act Waivers

Glenda L. Partee December 2012

Contents

- 1 Introduction and summary**
- 7 New evaluation systems and evidence of effectiveness**
- 11 What measures and what methods?**
- 27 What weights and what percentages?**
- 34 Making significant progress**
- 39 Findings and recommendations**
- 43 Conclusion**
- 44 About the author**
- 45 Endnotes**

Introduction and summary

Consensus is elusive when it comes to figuring out exactly what it takes to improve our nation’s public schools. When the quest is to ensure that our children achieve academically, there just aren’t many certainties. Except one: The quality of teaching matters.

Research shows that an effective teacher is key to student success. But determining what evidence best reflects teacher effectiveness and how this information can be used to improve the quality of teaching are among the significant issues facing public education today.

The impetus for meaningful teacher evaluation reform from many sectors set the stage for the major changes we are now witnessing in the direction and scope of teacher performance evaluation. Some of the factors leading to this reform include:

- The 2009 seminal report, “The Widget Effect,”¹ exposed the reigning indifference to instructional effectiveness in our schools and in our policies—an indifference that ignores variations in the effectiveness of our teachers, treating them as if they were all the same, and that does little to address the problem.
- Advocates are decrying the lack of state guidance and requirements for teacher evaluations. For too many school and district leaders, formal evaluation is a compliance activity instead of an opportunity to provide meaningful feedback to teachers for improvement.²
- Academics pronounce that the state of teacher performance evaluation is a non-system in need of major reform.³
- Many sectors—governors and mayors of different political parties, state legislatures, businesses, and educators and their unions—are calling for meaningful reforms in the way we evaluate and support our teachers.

Dynamic reforms effecting teacher evaluation and support are now happening in states and school districts. These reforms are inspired in part by the U.S. Department of Education's competitive grant programs, including Race to the Top, which require new standards and assessments in our public schools, data systems capable of measuring student growth, and human capital systems designed to recruit, develop, and retain effective teachers. This effort is matched by recent priorities of the Teacher Incentive Fund supporting district-wide evaluation systems that reward teacher success. The Education Department's decision to provide waivers from key provisions of or flexibility within the Elementary and Secondary Education Act—also known as No Child Left Behind—offers a further boost and a framework for states to make these long overdue reforms in a coherent way.

On February 28, 2012, 26 states and the District of Columbia submitted requests to the Department of Education for waivers. Twenty-three states were ultimately approved; two states (Idaho and Illinois) have pending applications; one state (Vermont) withdrew; and one state (Iowa) was rejected. (Note: Idaho's application was approved on October 17, 2012, while this paper was drafted and is therefore not a part of this analysis.) Eleven other states received waiver approvals in an earlier round.⁴ As part of the second round of requests, all states presented plans to raise standards, improve accountability, and support reforms to improve principal and teacher effectiveness. These plans provide an important view into the decisions and actions of states as they design, build on, or perfect the systems for these new reforms.

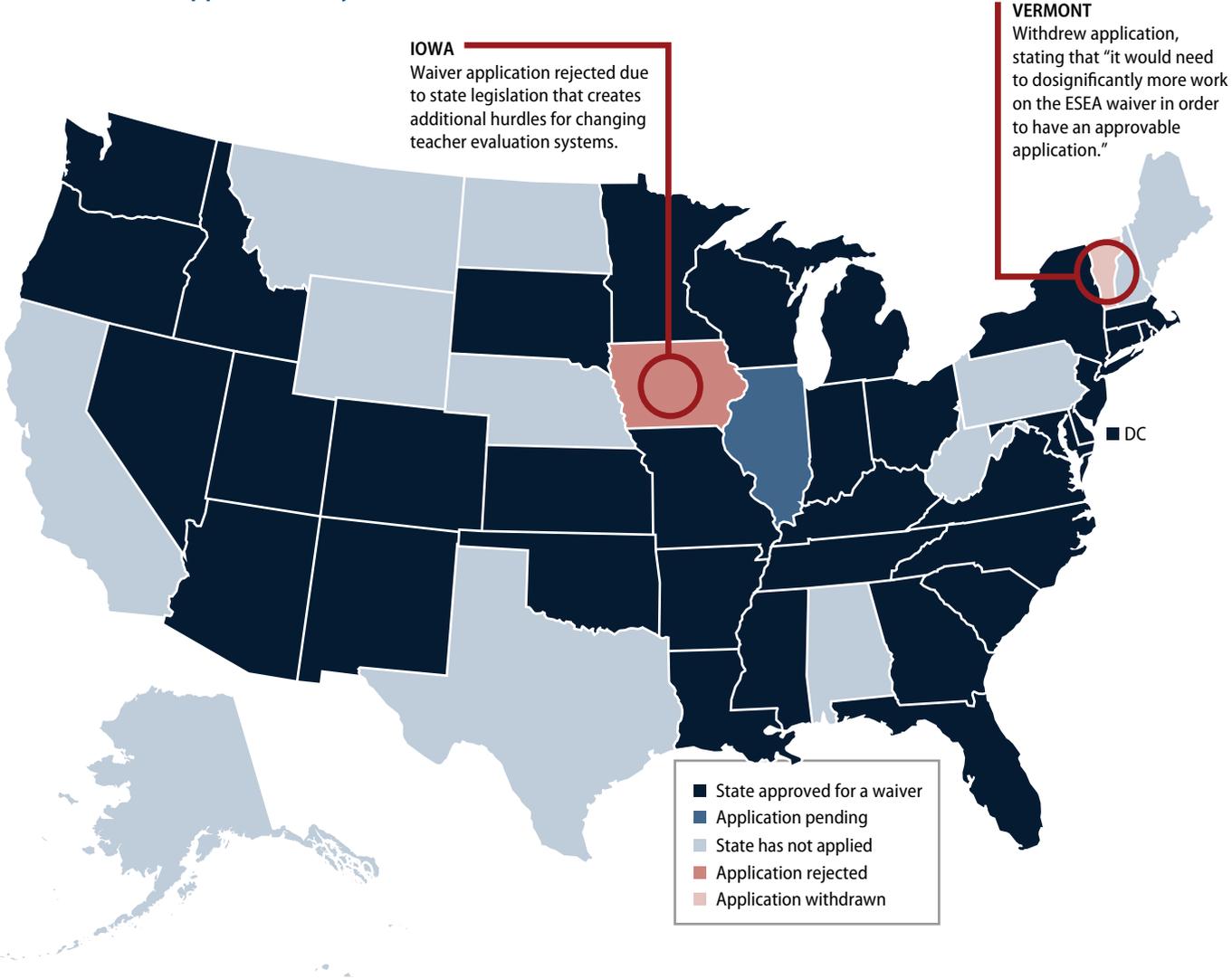
Many states are now actively building or implementing educator workforce systems with meaningful evaluation and support systems that are linked to improvements in classroom practices and student achievement. No longer is teacher evaluation expected to be merely perfunctory or used exclusively as the basis of personnel decisions. State leaders are rethinking the underlying assumptions and policies of teacher evaluation systems and, together with critical stakeholders, are planning the implementation of new systems.

The focus of this report is on one piece of this very large set of transformations: the multiple measures and multiple methods used in new teacher evaluation systems, including the weighting of these measures, to determine a composite score of teacher effectiveness. The data source for our analysis is the plans of 23 second-round waiver applicants approved by the U.S. Department of Education as of August 2012. These include the plans received and approved for Arizona,⁵ Arkansas,⁶ Connecticut,⁷ Delaware,⁸ the District of Columbia,⁹ Kansas,¹⁰

Louisiana,¹¹ Maryland,¹² Michigan,¹³ Mississippi,¹⁴ Missouri,¹⁵ Nevada,¹⁶ New York,¹⁷ North Carolina,¹⁸ Ohio,¹⁹ Oregon,²⁰ Rhode Island,²¹ South Carolina,²² South Dakota,²³ Utah,²⁴ Virginia,²⁵ Washington,²⁶ and Wisconsin.²⁷

Our review of these various reform plans indicates that the design and implementation of new systems of evaluation and support are truly works in progress. It's clear that this work will be an iterative process and that it should be open to review

FIGURE 1
Status of waiver applications, by state



Source: U.S. Department of Education, <http://www.ed.gov/esea/flexibility>
 Note: Idaho's application was approved on October 17, 2012, but it is not a part of this analysis.

and adjustment as new research and the results of pilot implementations surface. For now, the state efforts and the waiver process both represent a rich laboratory of exploration and reform that bears watching for lessons to be learned, as well as for necessary corrections to be made. A few findings have already emerged from this initial review. They include the following:

- **This is hard work that is being approached differently by states while they implement multiple reforms.**
 - It is difficult to legislate, regulate, and provide guidance for change within an environment of multiple simultaneous reforms. These reforms include the implementation of new college and career-ready standards, statewide data systems, new assessments, and new state responsibilities for these new systems, to name a few. The new educator evaluation systems must align with and be a part of these other reforms.
 - Each state approach, including that of the District of Columbia, is different, and each is at a different stage of development and implementation. Evaluation designs are influenced by factors such as the characteristics of local school districts, laws governing charter school autonomy, and a state’s history for local control and collective bargaining agreements related to educator evaluation.
- **Measures used to assess educator effectiveness are diverse and cannot be captured by only one or two indicators.**
 - Waiver winners rely on a range of measures and methods for assessing teacher professional practice, including classroom observations, self-assessments and reflection, teaching artifacts, student-learning measures, and surveys of students and parents.
 - States are using both student-achievement measures (measures of student learning at a specific point in time) and growth measures (changes in student learning over time), including value-added estimates based on state assessments when available, to capture measures of student success aligned with individual teachers or teams of teachers. Some states are still considering the types of student-growth measures to use, and some are piloting multiple models before recommending a particular approach.

- States are also looking to more personalized and school-appropriate measures for determining teacher impact on student learning and vesting teachers more directly in monitoring student progress through approaches such as student-achievement goal setting, student-learning objectives, student-learning targets, teacher goal setting, and unit work samples. These measures are used to actively engage the teacher and the evaluator in a goal-setting process for student learning that is customized for the teaching assignment and for the students.
- States give different weights to component measures devoted to indicators of student achievement and indicators of professional practice; they also rely on different measures. Some states have specific percentages of components spelled out in state law. Others do not. In some cases a certain amount of discretion is given to local districts for insertion of components they value in the evaluation.
- **States are expanding the measures used to determine teacher effectiveness for nontested grades and subjects.**
 - Though some states are in the beginning stages, all are determining or developing assessments applicable to teachers of grades and subjects that are not part of statewide, standardized assessments for the purpose of determining student growth.
 - Typically this involves expanding the portfolio of state assessments to provide growth data in all grades and subjects or expanding the portfolio of nationally or locally approved assessment tools that can be validly used such as classroom-based assessments, unit tests, end-of-course assessments, student-learning objectives, and portfolios.
- **Systems have diverse purposes.**
 - Waiver applicants were responsive to the application requirements making these systems as much about differentiating educators on their levels of effectiveness and for use in making personnel decisions as about using the evaluation process to identify areas for overall educator improvement.
- **Successful systems need an infrastructure of support.**
 - The work of the states is not just about creating new systems of teacher evaluation, but also about putting an infrastructure in place to ensure the success of

these systems. This means that teachers and principals must receive orientation to the new systems; evaluators must receive appropriate training (for example, in collecting evidence, rating against a professional standard, and providing feedback); rubrics and protocols for observation must be identified and tested; strong teacher-student data links must be in place that verify that the teacher of record is tied to the right students for purposes of assessing teacher impact; and management systems must be devised that allow teachers to track their progress toward learning goals. Just as importantly, supports and interventions must be in place to move teachers toward higher levels of effectiveness in line with the information provided through evaluation.

Against this evolving backdrop we offer the following policy recommendations:

- ♦ The U.S. Department of Education should closely monitor the successes and problems experienced by these states and the District of Columbia as they implement these new systems of evaluation and support them going forward.
- ♦ The states and the District of Columbia should continue to heed emerging findings from research and evaluation and seek feedback from their own efforts to ensure continuous improvements.
- ♦ The U.S. Department of Education and philanthropic organizations should continue to support improvements in the tools and infrastructure necessary for the development and sustainability of these new evaluation systems.
- ♦ Lessons learned from these efforts must inform the future direction of education reform through the reauthorization of the Elementary and Secondary Education Act.

New evaluation systems and evidence of effectiveness

Under the Elementary and Secondary Education Act waiver process, states are no longer required to submit highly qualified teacher improvement plans.²⁸ In exchange, state education agencies will agree to develop and adopt guidelines for local teacher and principal evaluation and support systems, and they will ensure that local education agencies implement these evaluation and support systems consistent with the guidelines of the state education agency. (See Figure 2 for criteria for approval of flexibility and Figure 3 for definition of student growth.)

FIGURE 2

Criteria for flexibility in supporting effective instruction and leadership

To receive this flexibility, an SEA [state education agency] and its LEAs [local education agencies] must commit to develop, adopt, and implement (with the involvement of teachers and principals) teacher and principal evaluation and support systems that:

- Will be used for continual improvement of instruction
- Meaningfully differentiate performance using at least three performance levels
- Use multiple valid measures in determining performance levels, including as a significant factor data on student growth for all students (including English Learners and students with disabilities), and other measures of professional practice (which may be gathered through multiple formats and sources, such as observations based on rigorous teacher performance standards, teacher portfolios, and student and parent surveys)
- Evaluate teachers and principals on a regular basis
- Provide clear, timely, and useful feedback, including feedback that identifies needs and guides professional development
- Will be used to inform personnel decisions

Note: The above information is quoted from: Department of Education, ESEA Flexibility: Frequently Asked Questions (2012), p. 31, available at <http://www2.ed.gov/policy/eseaflex/esea-flexibility-faqs.doc>

FIGURE 3
Defining student growth and achievement

“Student growth” is the change in student achievement for an individual student between two or more points in time. For the purpose of this definition, student achievement means:

- For grades and subjects in which assessments are required under ESEA [Elementary and Secondary Education Act] section 1111(b)(3): (1) a student’s score on such assessments and may include (2) other measures of student learning, such as those described in the second bullet, provided they are rigorous and comparable across schools within an LEA [local education agency].
- For grades and subjects in which assessments are not required under ESEA [Elementary and Secondary Education Act] section 1111(b)(3): alternative measures of student learning and performance such as student results on pre-tests, end-of-course tests, and objective performance-based assessments; student learning objectives; student performance on English language proficiency assessments; and other measures of student achievement that are rigorous and comparable across schools within an LEA [local education agency].

Note: The above information is quoted from: Department of Education, ESEA Flexibility (2012), p. 10, available at <http://www.ed.gov/eSEA/flexibility/documents/eSEA-flexibility.doc>.

This change in focus represents the insights gained since the implementation of No Child Left Behind in 2001—an important one being that attaining highly qualified teacher status is a minimum bar that varies from state to state and does not reflect teacher abilities to improve student learning.²⁹ This position reflects new findings from research and recent reforms in the states holding educators accountable for the success of their students, recognizing and rewarding educators for their effectiveness and, when necessary, dismissing those who are ineffective.

Some of this action has been prompted by competitive federal programs such as Race to the Top, which offered ample incentives for the states to improve teacher and principal effectiveness based on performance; to establish clear approaches to measuring student growth; to have local education agencies conduct annual educator evaluations; and to ensure the rigor of these evaluations, among other things. Forty-one states applied to the first round of the Race to the Top competition in January 2010 with proposals to implement these reforms. Other federal grant programs, such as the Teacher Incentive Fund, encouraged performance-based salary

approaches that drive the need for improved teacher evaluation systems. Federal policy continues to promote these types of reforms. The 2012 Teacher Incentive Fund competition included a new focus on supporting district-wide evaluation systems that reward success, offer greater professional opportunities, and drive decision making on recruitment, development, and retention of effective teachers and principals.

A number of states, even if they were not successful Race to the Top or Teacher Incentive Fund grantees, have made progress over the years in planning for and implementing many of the aforementioned reforms.³⁰ As a result, many were well positioned to accept the waiver challenge. Through the waiver process, and in the absence of a reauthorized Elementary and Secondary Education Act that would have captured these pivotal changes, state and local education agencies now have incentive to develop and implement more meaningful educator evaluation and support systems. Many states are well on their way to doing so. Many states have also bumped up against their 100 percent highly qualified teacher goals³¹ and recognize that this is at best a floor of expectation for teacher qualification that is limited by its focus on inputs to good teaching instead of the actual performance of teachers.³²

Guidance is now available from researchers and early implementers on the qualities of new and more rigorous approaches to evaluation. There is consensus that new evaluation systems must be based on fair and valid measures in order to adequately capture the complexity of good teaching and infuse more accuracy into the evaluation process, especially when this process is tied to high stakes personnel actions.³³ Multiple measures are needed to encompass the many purposes of a comprehensive approach that increasingly includes identifying teacher effectiveness, ensuring greater accountability for student learning, improving teacher practice by diagnosing areas in need of professional improvement and development, as well as determining personnel decisions.

As states build their new educator evaluation systems, they must make critical design decisions, including:

- Determining the right ingredients or valid measures necessary for creating a composite teacher rating that accurately reflects a teacher's effectiveness and can be used on a performance continuum
- Deciding what percentage of a teacher's total evaluation score should be linked to changes in student achievement or quantitative measures of student growth and what percentage should be allotted for qualitative measures of teacher practice

- Figuring out how these teacher evaluation results can best be used as the basis of personnel actions to support teacher professional growth and development, as a mechanism for aligning teacher and student effort and goals, or as a way of distributing strong educators equitably throughout a system, leveraging educator strengths, and allowing for differentiated job responsibilities
- Determining to what extent districts should have flexibility in the identification, use, and weighting of evaluation components

States are fully engaged in making these important decisions. For the waiver applicants, the process of determining the quantitative and qualitative multiple measures and methods to be used has been lengthy, difficult, and, in some states, contentious. There are lots of moving parts and in some cases the policy decisions have gotten ahead of the tools of evaluation, but improvements continue and the field gets smarter. For these reasons, this will likely be an iterative process and should be open to review and adjustment. The design and implementation of these systems will not be perfect in their first or second iterations. For now the state efforts and the waiver process represent a rich laboratory of exploration and reform that merit watching, both for lessons to be learned and for necessary corrections to be made.

What measures and what methods?

According to the Department of Education's guidance for waiver flexibility,³⁴ measures used in the performance evaluation systems must be clearly related to increasing student academic achievement and school performance. The Department further asks:

*Does the SEA incorporate student growth into its performance-level definitions with sufficient weighting to ensure that performance levels will differentiate among teachers and principals who have made significantly different contributions to student growth or closing achievement gaps?*³⁵

Reflecting the guidance from the Department of Education, the bases for teacher evaluations used by the waiver applicants are typically divided into (1) measures of professional teaching practice, though in some cases this category is split by the states to also represent professional responsibilities, and (2) measures of student achievement.

Measures of professional practice

Experts stress that the qualitative measures used to determine instructional quality or professional practice must be founded on high-quality standards of what is known about effective teaching practices. These standards must be clear and transparent about what effective teaching practice looks like.³⁶ While there are no national standards, some states have adopted or use some variation of the Council of Chief State School Officers Interstate Teacher Assessment and Support Consortium, *Model Core Teaching Standards*³⁷ (for example, Arizona, Mississippi, Utah, South Carolina, Virginia, and Wisconsin), and/or the National Board of Professional Teaching Standards (for example, Mississippi and Virginia). Other states have created their own standards based on research and stakeholder input (for example, Connecticut, the District of Columbia, Missouri, Nevada, New York, Ohio, and Rhode Island).

Once standards are determined, professional practice may be assessed using a combination of the following:

- Observations, including feedback from peers, based on rubrics aligned with standards of professional practice. Many states are using the Charlotte Danielson *Framework for Teaching*³⁸ as their evaluation rubric for assessing educator practice. These states are Alaska, Delaware, Mississippi, Louisiana, Maryland, South Carolina, South Dakota, and Wisconsin.
- Self-assessments and reflection.
- Artifacts—or documents that reflect some aspect of classroom teaching that is not directly reflected in classroom practice—such as lesson plans, unit work samples, curriculum design, pacing guides aligned with the standards, student assignments, portfolios, and evidence of field experience.
- Student-learning measures such as samples of student work, including portfolios and research papers.
- Student and parent surveys.

How these measures are combined can be seen in the six measures used in South Carolina's Assisting, Developing and Evaluating Professional Teaching system to determine teacher performance levels. These measures include:

- Teachers' long-term plan(s)
- Classroom observations, with a minimum of four unannounced visits per year and additional walk-through observations permitted
- Teacher reflections following each classroom observation
- Professional performance review completed by the principal (or designee and other supervisors)
- Professional assessment completed by the teacher, which is the first step to developing the teacher's professional growth and development plan
- One or more unit work samples (a demonstration of student learning which is discussed later in this paper)

Observations: Classroom observations have traditionally been the staple of teacher evaluations, though they are often derided as perfunctory and as failing to clearly distinguish levels of performance. Experts suggest that to be useful, observations must adhere to research-based rubrics that distinguish performance levels (at least four³⁹), define behaviors and practices of excellent educators, differentiate between veteran and novice educators, and provide a roadmap for improvement.⁴⁰ Useful observations are valid, that is, they focus on behaviors that matter for student learning, are reliable, and require an appropriate infrastructure for successful implementation (such as trained observers, valid rubrics, and formal protocols). For these reasons, caution is counseled in the purposes and uses of these measures. Appropriate purposes include identifying individual or programmatic areas of strength and areas in need of improvement and determining individualized professional development and support.⁴¹ Because of these concerns, multiple (not single) observations should be among the several measures used in a comprehensive teacher evaluation system.

Emerging research underscores this point. The findings from the Measures of Effective Teaching project provides extensive guidance to policymakers and practitioners on improving teaching and learning through better evaluation, feedback, and professional development.⁴² Research on the value of five classroom observations tools found them positively associated with student achievement gains,⁴³ but to reliably characterize a teacher's practice requires averaging scores over multiple measures. Combining observation scores with

evidence of student achievement gains and student feedback also improved predictive power and reliability. Finally, the combined measure identifies teachers with larger gains on state tests of student achievement than traditional measures of teacher experience and graduate degrees. Teachers with strong performance on the combined measure also performed well on other student outcomes.⁴⁴

Peer review: Another form of observation is peer review, which can be used to provide feedback on instruction in a formative manner or as part of a formal summative review. Peer review is often a collaborative process in which the teacher works closely with a colleague or a group of colleagues to improve instructional strategies.⁴⁵ It is seen as a way of empowering teachers in the evaluation process. Known as peer assistance and review, this approach uses senior teachers to mentor both newcomers and struggling veteran teachers, and it is considered a strong form of professional development, although an outcome of a peer review can be teacher dismissal.

Student surveys: Researchers increasingly believe that student surveys can provide important insights into a teacher's effectiveness. This measure is among those studied in the Measures of Effective Teaching project, which found that student feedback was a better predictor of a teacher's performance than more traditional indicators of success such as whether a teacher had a master's degree. The Tripod Survey, a reliable measure and predictor of student achievement gains, is used to gauge seven areas of classroom life and teaching practices and is either in use or under consideration in a number of waiver states.⁴⁶

Combinations used by selected waiver applicants are described below.

- **Arkansas:** The state determines qualities of teaching through observation rubrics and artifacts such as lesson plans or pacing guides aligned to the state standards. Other measures include self-directed or collaborative research approved by the evaluator.
- **Arizona:** The state allots 50 percent to 67 percent of an evaluation total for evidence of teaching performance. The protocol for evidence requires that it provides for periodic multiple observations of all teachers.

- **Connecticut:** Along with teacher observation and professional practice, which accounts for 40 percent of the total evaluation, the state uses feedback from peers and parents, including surveys (10 percent) and schoolwide student-learning indicators or student feedback (5 percent).
- **Delaware:** The state supports continual improvement of instruction through rubrics based on Charlotte Danielson’s framework to assess: planning and preparation; classroom environment; instruction; and professional responsibilities. The fifth component is tied to student improvement (growth) measures and becomes the gatekeeper; to be rated “effective,” an educator must demonstrate “satisfactory” levels of student growth. Of the components, the fifth must be weighted as highly as any other component.
- **District of Columbia:** Instructional expertise at District of Columbia Public Schools is based on up to five formal observations each year, three by administrators and two by independent expert master educators,⁴⁷ as well as measures of teacher collaboration and professionalism.
- **Kansas:** The state is in the early stages of determining its measures and developing and adopting guidelines, and is conducting pilot studies of artifacts that impact student achievement. Among the measures under review are observations, including those by peers, professional growth, self-reflection, student voice, parent voice, and others.
- **Maryland:** Fifty percent of the state’s evaluation model must allow for professional practice based on the four components of the Danielson framework. In addition to these four qualitative measures, local education agencies can include other local priorities on which they may want to hold teachers responsible.
- **Nevada:** Teacher performance based on a self-assessment of high-leverage instructional principles, as well as professional responsibilities, will account for 50 percent of teacher evaluation results, although specific indicators are in development. The evaluation process will include a self-assessment; a pre-evaluation conference between teacher and evaluator; the announced observation (the number of which will be based on whether the teacher is probationary, or is deemed ineffective, minimally effective, effective, or highly effective); and a post-action review that includes standardized questions and potential artifacts/evidence requested by the evaluator. Year-to-year student outcome data are also part of the evaluation cycle and are used to guide professional development decisions.

- **New York:** The state uses an evaluation rubric aligned with relevant standards that includes multiple classroom observations. The rubric can include other methods, as well, such as observations by independent evaluators, state-approved surveys of students and parents, or structured reviews of teacher artifacts of practice.
- **Oregon:** Its evidence of professional practice includes assessment through classroom observation and examination of artifacts. Peer evaluation is encouraged but can only be used in the formative evaluation process in order to identify educator strengths and weaknesses during the instructional process, not as a measure of summative evaluation, which is used to determine the educator's ultimate effectiveness.
- **Rhode Island:** Rhode Island requires at minimum both formal and informal observations of educator practice using valid and accurate rubrics and tools. The evaluation rubrics are designed to facilitate constructive and timely feedback, which leads to the development of individualized professional development plans. Evaluation systems must also include information from students' parents, assessments of professional responsibilities, and areas of practice and student learning.
- **South Dakota:** Fifty-percent of the teacher evaluation is to be based on observable, evidence-based characteristics of good teaching and classroom practices. Districts may collect additional evidence through, for example, classroom drop-ins, peer review, parent surveys, student surveys, or portfolios.
- **Utah:** Observations of instructional quality are to account for a minimum of 40 percent of the overall evaluation score. Parent and student input measures are pending based on the results of pilot studies, but they likely won't account for more than 20 percent.
- **Virginia:** The state's *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers* includes seven performance standards, with the first six encompassing measures of teacher practice: professional knowledge; instructional planning; instructional delivery; assessment of and for student learning; learning environment; and professionalism. The seventh, student academic progress, is discussed in the following section.
- **Washington:** School districts are to include unobservable evidence of practice such as artifacts, as well as observation and observable evidence of practice.

(Examples of evidence of practice are on page 13.) Districts may use classroom-based, school-based, district-based, and state-based tools, all of which may include perceptual data from students.

Alaska, Nevada, Oregon, and Rhode Island also include evidence of professional responsibilities, which focuses on the role and responsibilities of the teacher within the learning community and the contribution of teachers to school-wide goals. Measures of professional responsibilities used include: self-reflections and reports; professional goal setting; student-growth goal setting; peer collaboration and teamwork; records of contributions such as the building-level leadership, participation on committees, and the meeting of professional obligations; and family engagement strategies.

Measures of student achievement and growth

In addition to measures of professional practice, waiver winners are using both student achievement measures (measures of student learning at one point in time) and growth measures (changes in student learning over time) where available. Michigan, Nevada, Utah, and Wisconsin are still considering the types of student-growth measures to use; other states are piloting multiple models before they recommend a particular approach. All states—though some are in the beginning stages—are determining or developing assessments applicable to teachers of grades and subjects that are not part of statewide standardized assessments for the purpose of determining student growth.

Whereas growth measures tied to national or state assessments are used in evaluations to assess teacher impact on student learning, states are also looking to more personalized and school-appropriate measures for determining teacher impact on student learning and vesting teachers more directly in monitoring student progress. Whether called student-achievement goal setting (Virginia),⁴⁸ student-learning objectives (Connecticut, Maryland, Missouri, Oregon, Ohio, Rhode Island, Utah, and Wisconsin), student-learning targets (Louisiana), teacher goal setting (Oregon), or unit work samples (South Carolina), these measures are used to actively engage the teacher and the evaluator in a goal-setting process for student learning that is customized for the teaching assignment and for the students. These measures are often used in addition to valid external measures of student academic progress or when these other measures aren't available.

FIGURE 4
Rhode Island's student-learning objectives

Student-learning objectives are not student specific. Rather, they are long-term academic growth targets that a teacher sets for students or subgroups of students within a classroom. The development of student-learning objectives requires close collaboration of teachers and their school administrator to determine clear expectations for student learning, learning targets, and how learning should be assessed. Teachers actively use data to set measureable targets for how much their students will learn over the course of instruction, and they must closely monitor student progress. According to the Rhode Island waiver application:

A Student Learning Objective is a long-term (typically one semester or one school year) academic goal that teachers set for groups of students. It must be specific, measureable, based on available prior student-learning data, and aligned with state standards as well as with relevant school and district priorities. ... All teachers of the same course in the same school use the same set of objectives, although specific targets may vary if student starting points differ among classes.

Source: U.S. Department of Education, Rhode Island ESEA Flexibility Request (2012), pp. 119–120.

Let's examine which measures of student achievement and growth are in use in the selected states and the District of Columbia, as well as other evaluation measures these states are employing. This information illustrates the diversity and complexity in how the states and the District of Columbia are approaching their charge.

Evidence of student growth as a significant factor

Measures of student growth are developed using student test scores from two or more years and focus on performance of individual students. Results of these measures indicate whether a student is on track to reach a proficiency performance level. Growth models are important because, conceptually, they align well with student learning, provide richer information on student learning than any single test score, and focus on the development of individual students.⁴⁹

Teacher value-added models are the most sophisticated of the test-based growth models and attempt to measure what educators contribute to their students' test scores. Whereas student scores are the sole variables used in student-growth models, value-added models use student scores along with student and teacher variables.⁵⁰

The use of value-added models is relatively new and controversial. Among its limitations are that strong student-teacher links are not always available, and value-added estimates can only be calculated for teachers of tested grades and subjects.⁵¹ As far as criticisms go, researchers have found that value-added models of teacher effectiveness do not produce stable ratings of teachers, and that evaluation scores can fluctuate from class to class and year to year. Moreover, even under the best circumstances, a teacher's efforts represent just one element of many conditions impacting student success.⁵² Despite the shortcomings, value-added measures are useful, especially when they are combined with other measures. This results in a more complete picture of teacher effectiveness.⁵³ Value-added measures show positive relationships to other teacher performance measures such as classroom observations and principal evaluations.⁵⁴

State standardized assessment tests are the most frequently used external measures for student growth, but the results only apply to teachers of tested grades and subjects. Arkansas illustrates the complexity of this issue: Summary growth statistics are available at the teacher level for grades four through eight in math and literacy, and median summary growth percentages are available for grades one through nine in reading and math; grades three through eight in math and literacy; grade five and grade seven in science; grade 11 in literacy; and for end-of-course exams in algebra, geometry, and biology in whatever grade they are taken. There is currently no consensus regarding the appropriate growth measures to incorporate in Arkansas's evaluation system. In order to keep its options open for transition to the new Partnership for Assessment Readiness for College and Careers assessments,⁵⁵ modeling of student achievement and growth at various weights will be incorporated into Arkansas's 2012–13 pilot implementations using growth to standard and student-growth percentile models.

States use a range of “other” student achievement measures

To address the limitation to growth measures presented by teachers of subjects and in grades without comparable growth measures, states use a variety of other measures of student achievement:

- **Arkansas:** Measures include classroom assessments such as samples of student work; portfolios; writing projects; unit tests; pre- and post-assessments; classroom-based formative assessments; district-level assessments, including formative assessments, grade- or subject-level assessments, department-level assessments, and common assessments; state-level assessments, including end-of-course assessments, statewide assessments of student achievement, and career and technical assessments; and national assessments, such as the Advanced Placement program.
- **Arizona:** The state’s evaluation framework provides for a sliding weight across three components with 1) 33 percent to 50 percent tied to student quantitative data such as the Arizona state assessment; Stanford 10; Advanced Placement; International Baccalaureate; ACT (formerly American College Testing); or district- and charter-wide assessments; 2) an optional 17 percent tied to school-level and/or system-level data; and 3) 50 percent to 67 percent reflecting professional practice. This sliding framework is designed to provide local education agencies with maximum flexibility while at the same time recognizing the different assessment data available across different grades and content areas.
- **Connecticut:** Student-learning indicators must account for 45 percent of the evaluation, with half of that based on the state test for tested grades and subjects, or another standardized assessment for grades and subjects for which there is no state test. The other half comes from examples of student-learning indicators, including teacher-developed assessments, portfolios of student work, and student-learning objectives.
- **Delaware:** The state uses a student-growth model for teachers in the tested subjects and grades. For other teachers, external measures (such as SAT, ACT, or Star Reading) have been identified and are under review for validity, reliability, and rigor. Additional internal measures (aligned with specific state standards and correlated with class instruction) are being developed by educators across the state and will be rolled out for use by local education agencies for various cohorts of teachers.⁵⁶ Delaware expects to have full multiple measures identified and approved for all teachers, specialists, and administrators for the 2012–13 school year, as well as a fully implemented system. Of the five-component evaluation measures in the state, one is devoted to student growth and can only be weighted as high as any other component.

- **District of Columbia:** The design of the District of Columbia’s evaluation systems is driven by whether local education agencies (such as charters and the District of Columbia Public Schools, which is the largest local education agency) participate in its Race to the Top grant and the dictates of its charter school law. District of Columbia Public Schools must include student achievement for 50 percent of teacher evaluations in tested grades and subjects. Specifically, District of Columbia Public Schools will include a growth measure based on the state test for at least 30 percent of the evaluation rating and may select another measure of achievement or growth for up to 20 percent of the evaluation rating. For teachers in nontested grades and subjects in District of Columbia Public Schools, a measure of growth will account for at least 15 percent of the evaluating rating. Charter Race to the Top local education agencies will be required to use the District’s value-added model as 50 percent of the evaluation rating for teachers in tested grades and subjects unless the local education agency receives a waiver from the state office. Charter local education agencies with waivers will have flexibility in the weights assigned to student-growth measures for teachers in nontested grades and subjects. If a charter’s waiver is approved, it must use the value-added model for at least 30 percent of the rating and can propose other measures of achievement for the remaining percentage to equal 50 percent.
- **Louisiana:** Beginning in the 2012–13 school year, all educators in the state will be evaluated annually, including those in nontested grades and subjects, with 50 percent of the evaluation based on measures of student growth and 50 percent based on observation and other measures of effectiveness. A statistical co-variate value-added model that controls for prior student achievement and other variables will be used for tested grades and subjects, but the number of value-added measures is expanding through adoption of valid state assessments for more subjects and grades. In the meantime, valid state-approved common assessments—such as Advanced Placement exams or the Developmental Skills Checklist for kindergarten readiness—can be used as measures of student growth along with rigorous student-learning targets for nontested grades and subjects. These are comparable to the student-learning objectives discussed earlier.
- **Maryland:** The state is in the process of developing model evaluation criteria to measure state performance on student growth. This will account for 50 percent of a teacher’s or a principal’s evaluation. Student growth will be determined based on the courses and grade levels that a teacher teaches. The state model

also incorporates the Maryland School Performance Index⁵⁷ and student-learning objectives to define student growth. Where a statewide assessment exists, it must be used as one of the multiple measures. State assessments, if available, will be combined with student-learning objectives at the state education agency's approval to yield teacher ratings.⁵⁸

- **Michigan:** The state is in the process of developing recommendations for its statewide evaluation system. By statute, however, it will include a statewide student-growth and assessment tool for use by all content areas and measure growth for students at all achievement levels. The plan is to expand the portfolio of state assessments—or expand the portfolio of approved national or local assessment tools—to determine growth in all grades and subjects. State legislation requires 25 percent of educator evaluations to be based on student-growth and assessment data by the 2013–14 school year, 40 percent of educator evaluations by the 2014–15 school year; and 50 percent of educator evaluations by the 2015–16 school year.
- **Mississippi:** The state's teacher appraisal guidelines are currently in the pilot phase, and a protocol to measure student growth that can be linked to teacher performance is under development. For teachers in nontested grades and subjects, student progress will be determined by student-growth percentiles on statewide assessments at the school-wide, not the teacher, level.
- **Missouri:** The state is conducting a student-growth pilot project in 156 districts focusing on student growth and value-added measures. Findings from these two models will inform the state's evaluation guidelines and its model evaluation system. For nontested grades and subjects, district-generated assessments, student-learning objectives, and results of end-of-course tests are among potential evidence of student achievement that will be included in the model system. Professional impact on student learning is one of three frames in Missouri's educational evaluation system. The other two frames are professional commitment and professional practice.
- **Nevada:** By state statute, evaluations using multiple methods are to be based at least on 50 percent for student outcomes. Under draft guidelines, an index for student outcomes will include student growth (accounting for 20 percent), student proficiency (accounting for 15 percent), teacher contributions to reduction in subpopulation gaps (10 percent), and student engagement based on the Tripod Survey (5 percent). For teachers in grades and subjects where statewide

assessment data do not exist, the state board of education will regulate measures that local education agencies may use to determine student growth. For now Nevada is looking to districts with federal School Improvement Grants and Teacher Incentive Fund support that are using aggregate or school-wide data to generate shared attribution scores for teachers at the school level. Validation and pilot efforts of potential solutions for student-growth measures for all teachers will extend through the 2013–14 school year.

- **New York:** Student achievement measures in New York account for 40 percent of the composite effectiveness score, with 20 percent based on student growth on either the state assessments or other comparable measures where state assessments are not available. This increases to 25 percent when the value-added growth model is implemented in the 2012–13 school year. An additional 20 percent is based on valid and reliable locally selected measures of student achievement.⁵⁹ (This decreases to 15 percent when the value-added model is implemented.) New York plans to extend its growth/value-added model to its high school Regents exams. It also expects to add exams for additional subjects such as middle school science and social studies and high school English so that the growth model impacts at least 50 percent of teachers.
- **North Carolina:** In its Race to the Top application, North Carolina committed to the inclusion of student growth in teacher evaluation instruments. Teacher contribution to student academic success is now one of six standards on which teachers are evaluated. Three methods will be used to determine a teacher’s individual growth value: (1) analysis of student work (used with grades and courses that focus on performance standards); (2) pre-post test growth model (used with grades and courses with statewide assessments, but where the Education Value-Added Assessment System cannot be used, for example in the early grades); and (3) the Education Value-Added Assessment System model (where there are statewide assessments and a prediction model has been determined).⁶⁰ The state board of education will establish permanent components of the sixth standard rating and their respective weights in 2012-2013.⁶¹

North Carolina already administers a number of statewide standardized assessments; these align with 40 percent of the teacher workforce.⁶² For remaining nontested grades and subjects, teacher design groups are creating other measures to assess student learning based on the Common Core State Standards, the North Carolina Essential Standards, the Occupational Course of Study, and the Extended Content Standards for Exceptional Children. Additionally,

a “team” growth value—on groups of teachers who share instructional responsibility for students—was piloted in 28 school districts during spring 2012, . These same school districts are also piloting the Cambridge Education Tripod Project student surveys. Depending on the outcomes of these pilots, a team growth value and student survey results will also become parts of the student-growth component where appropriate, as will the individual and school-wide growth values beginning in the 2012–13 school year.⁶³

- **Ohio:** Student value-added measures account for 50 percent of the state’s composite evaluation. The list of assessments that may be used to measure student growth when value-added measures are not applicable in nontested subjects and grades has not been finalized. The assumption is that a growth model will support teachers in core and noncore content areas and grade levels including pre-K through grade two; English language acquisition; music and physical education, and teachers who work with students with disabilities and gifted students. Ohio is designing guidance and resources for measuring growth in nontested subjects and grades, including end-of-course exams and student-learning objectives. All teachers will have one or more measures of student growth from the following categories: value-added scores, assessments from a state-approved list, and locally determined measures (such as student-learning objectives and shared attribution measures to encourage collaborative goals). The latter may include building-level or district-wide value-added scores or composite value-added scores for building teams such as content area, performance index gains, and building- or district-based student-learning objectives.
- **Oregon:** The state will use a teacher goal-setting approach to assess student growth. Teachers, in collaboration with their supervisor/evaluator, will be required to establish at least two student-learning goals (aligned to standards the teacher is expected to teach and students are expected to learn) and identify strategies and measures to be used to determine goal attainment. Teachers who are responsible for student learning in tested subjects and grades (English language arts and math in grades three through eight, and 11th grade) will use assessments from Category 1 as one measure. Category 1 includes state or national assessments such as Oregon Assessment of Knowledge and Skills, SMARTER Balanced, English Language Proficiency Assessment, or Extended Assessments. Teachers will also select one or more additional measures from Category 2 (common national, international, regional, and/or district-developed measures such as ACT, PLAN, EXPLORE, AP, IB, and Dynamic Indicators of Basic Early Literacy Skills, or others approved by the district or

state) or Category 3 (classroom-based or school-wide measures such as student performances, portfolios, work samples, and tests). Teachers in nontested grades and subjects will use measures that are valid representations of student learning from at least two of the three categories as appropriate.

- **Rhode Island:** Rhode Island requires the most heavily weighted component of evaluations to be based on evidence of impact on student growth and academic achievement. The Rhode Island growth model will be used to measure student learning for teachers in state-tested grades (third through seventh grade for English language arts and math). To ensure that growth in student learning is assessed in every classroom, grade, and course, student-learning objectives will also be used statewide.
- **South Carolina:** Student growth is being added as a new component of South Carolina's teacher evaluation system. An educator-evaluation stakeholder group is considering types of potential growth measures, including value-added models, unit work sample rating, school-level rating, common assessments, projects, and assignments. The South Carolina department of education is looking at the 59 schools that currently participate in the state's Teacher Advancement Program to serve as incubators for value-added assessments for teachers in tested subject areas and grades. For all teachers, including those in nontested subject and grades, the unit work sample process is being considered to provide student-growth data.⁶⁴ The weighted values of these measures have yet to be determined.
- **South Dakota:** The state is developing administrative rules for the specifics of its statewide evaluation system. By law, however, 50 percent of the teacher evaluation must be based on quantitative measures of student growth, which must in turn be based on a single year or multiple years of data from state validated assessments. For those teachers in grades and subjects for which there is no state assessment, success in improving student growth can be demonstrated using objective measures, which can include portfolio assessments, end-of-course exams, and other district approved assessments.
- **Utah:** The state will consider both achievement and growth measures for tested and nontested subjects. State board of education rule R277-531-3 requires every local education agency evaluation system to include valid and reliable measurement tools, including at a minimum, observations of instructional quality (to account for at least 40 percent of the overall score) and evidence of student growth. While the weighting is under consideration pending piloting and valida-

tion of the measure, a floor of 40 percent of the overall weighting for student growth will be used as a target.

Student-growth measures are to be phased in starting with the 2013–14 school year. For tested subjects, end-of-level tests are under development to align with the Utah Core Standards. Utah has chosen the value-added model of student-growth percentiles. Nontested subjects will be aligned with student-learning objectives currently under development. Whether teachers are linked to tested or nontested subjects, they will be required to develop student-learning objectives and be linked to growth in both areas. In addition, formative, interim, and summative assessments are being developed to provide student achievement data.

- **Virginia:** The state uses student-growth percentiles based on state tests. Where student-growth percentile data are not available or are inappropriate, districts must first look to validated quantitative measures of student academic progress that are already in use locally. Other measures can be used when two valid measures of student academic progress are not available, including the student achievement goal setting described earlier, among others. Student academic progress is to account for 40 percent of the teacher’s summative evaluation, of which 20 percent is based on student growth and the other 20 percent is based on one or more alternative measures.
- **Washington:** Measures of student growth are among three sources of evidence of teacher effectiveness used by the state, though the specific percentage to be attributed to student growth has yet to be determined.
- **Wisconsin:** The state’s measures of student achievement will comprise 50 percent of the overall evaluation system. Although all teacher evaluations will be based on multiple measures of student outcomes, the measures used and their relative weights will vary based on availability of measures. A growth score, for example, cannot be calculated at the high-school level because the state assessment is administered only once in high school. The weights will therefore look different by school level. There has been no consensus on a particular value-added growth model. The state department of public instruction is currently monitoring multiple models (value-added models and student-growth percentiles). These determinations will be made prior to a full piloting of the evaluation model during the 2013–14 school year.

When results from state assessments (producing value-added data for tested grades and subjects), district assessments, and student-learning objectives are available, equal weight will be given to these three measures. When only two of these measures are available, equal weight will be given to those two measures. When only student-learning objectives are available, they will account for 45 percent of the overall rating. In all cases, district improvement strategies and school-wide data will together comprise 5 percent of the student achievement data.

Measures to be used for teachers of covered grades and subjects are to include the following: individual value-added data (currently only possible for grades three through seven in reading and mathematics); district-adopted standardized assessment results informed by district and school goals, the Common Core State Standards, and 21st Century Skills; student-learning objectives agreed upon by teachers and administrators; and district choice of data based on improvement strategies and aligned to school and district goals. Measures for teachers of noncovered grades and subjects will include everything above except the value-added data.

What weights and what percentages?

States give different weights to component measures devoted to indicators of student achievement and growth and indicators of professional practice; they also use different measures. Weights range from as low as 20 percent to as high as 50 percent devoted to student-growth measures, with the remainder devoted to measures of professional practice.

As discussed earlier, some states are still developing the components of their systems, including devising guidelines and/or modeling various components before settling on required or recommended tools, methods, components, and weights for these components. Some states presented their applications in the midst of proposed regulation changes that would support their evaluation designs. Some states have specific percentages of components spelled out in state law, while others do not have specific percentages. In some cases discretion is given to local districts.

As part of the waiver review, the U.S. Department of Education asks whether the state education agency incorporates student growth into its performance-level definitions with sufficient weighting to ensure that performance levels will differentiate among teachers who have made significantly different contributions to student growth or to closing achievement gaps. For this verdict, the jury is still out.

In fact, some researchers warn against the precipitous weighting of mandated components such as student growth in state law until all the properties of the other components (their reliability and validity) are known and assessed. According to Matthew Di Carlo, senior fellow at the nonprofit Albert Shanker Institute, the manner in which these components add up to a teacher's total score is as important as the properties of any individual component. Only with both can one begin to assemble the right components and weight them accordingly into a composite teacher rating.⁶⁵ Perhaps it is for these reasons that some states are still in deliberations about the weight of components, although a number of states—

particularly the Race to the Top winners—have been at this work for a longer time and are therefore more definitive in their weighting of components.

Although the discussion below describes the various ways in which states have allotted or plan to allot the percentage of the total evaluation based on student performance data and evidence of professional practice, the picture is much more complex than this simple dichotomy suggests. The full story of multiple measures and methods used in the new evaluations resides in the details of their various approaches, but these are beyond the scope of this paper.

Table I attempts to capture these differences. It lays out the states by the percentage of their evaluation that is tied to student-performance data (the first percentage referenced) and by the professional practice indicators shown in the subsequent percentages.

TABLE 1
Percentage distribution by evaluation components: Second round approved waiver states

States	Comments	50/50	50/50 (roughly)	40/60 (roughly)	Equal across multiple categories	In development	No specific percentage
DC	Only applies to Race to the Top local education agencies. D.C. Public Schools to include a 50 percent student achievement measure (includes a growth measure on the state test for at least 30 percent and may include another measure of achievement or growth up to 20 percent) for teachers in tested grades/subjects. For teachers in non-tested grades/subjects, growth measure will account for at least 15 percent of the rating. Charter RTTT local education agencies will use the D.C. value-added model at 50 percent for teachers in tested grades/subjects unless they receive a waiver from the state education agency.	√					
LA	Louisiana Act 54 requires 50 percent based on measures of student growth, including non-tested grades and subjects, or NTGS; and 50 percent based on observations and other measures of effectiveness beginning in 2012-2013. The average of the two determines the overall composite score, which will translate into the overall effectiveness rating.	√					

States	Comments	50/50	50/50 (roughly)	40/60 (roughly)	Equal across multiple categories	In development	No specific percentage
MD	Developing model state performance evaluation criteria for student growth that accounts for 50 percent of a teacher's evaluation; and for professional practice that accounts for an equal 50 percent. Professional practice includes four qualitative measures based on the Danielson Framework.	√					
MI	Though the work is still in progress, state legislation requires the following: by 2013-2014, 25 percent of the annual year-end evaluation based on student growth and assessment data; by 2014-2015, 40 percent of the annual year-end evaluation based on student growth and assessment data; and by 2015-2016, 50 percent of the annual year-end evaluation based on student growth and assessment data.	√ Eventually					
MS	The Mississippi Teacher Appraisal guidelines are currently in the pilot phase. Measures of effectiveness to be used include 50 percent based on student growth; and an additional 50 percent based on a combination of teacher actions, in turn based on the Danielson Framework (30 percent), and Professional Growth Goals (20 percent).	√					
OH	Student value-added measures account for 50 percent; teacher performance measures account for 50 percent and are based on the seven Ohio Standards for the Teaching Profession.	√					
NV	By statute, evaluations are to be based at least 50 percent on student outcomes, including student growth and other measures; and 50 percent on measures of teacher performance, including instructional practice and professional responsibilities. Under draft guidelines, an index for student outcomes will include student growth (20 percent), student proficiency (15 percent), contributions to reduction in subpopulation gaps (10 percent), and student engagement (5 percent).	√					
SD	By law (HB 1234), 50 percent of a teacher's rating will be based on quantitative measures of student growth, and 50 percent will be based on qualitative evidence-based characteristics of good teaching and classroom practice. School districts may collect additional qualitative evidence. Administrative rules for the specifics are under development.	√					

States	Comments	50/50	50/50 (roughly)	40/60 (roughly)	Equal across multiple categories	In development	No specific percentage
WI	The Wisconsin Framework for Educator Effectiveness measures of student achievement comprise 50 percent of the overall evaluation system. Measures of educator practice account for 50 percent and are based on the InTASC standards and the Danielson Framework.	√					
AZ	The model framework sets guidelines for three required components of which 33 percent to 50 percent must be tied to student quantitative data; an optional 17 percent can be tied to school-level and/or system-level data; and 50 percent to 67 percent must be aligned to teaching performance reflective of the InTASC teaching standards.		√				
CT	Specifies a framework that includes the following: 45 percent, half of which is based on the state test for tested grades/subjects or other standardized assessment for those grades and subjects for which there is no state test, and the remainder on other student-learning indicators (teacher-developed assessments, portfolios of student work, and student-learning objectives); 40 percent on teacher observation and professional practice; 10 percent on feedback from peers and parents; and 5 percent from school-wide student-learning indicators or student feedback.				√ (45/40/10/5)		
UT	Weighting of student growth measures is under development pending piloting and validation. For now, a floor of 40 percent of the overall weighting for student growth will be used as a target. Observations of instructional quality are to account for at least 40 percent of overall score at minimum. Parent/student inputs are also to be determined pending piloting but likely for no more than 20 percent.				√ 40/40/20 under consideration		
NY	Twenty percent based on student growth on state assessments or on other comparable measures of student growth if such growth data are not available (increased to 25 percent upon implementation of a value-added growth model in 2012-2013); and 20 percent based on locally selected measures of student achievement (decreased to 15 percent upon implementation of a value-added growth model in 2012-2013). Sixty percent using an evaluation rubric aligned with the relevant standards, and including multiple classroom observations. This can also include other measure approaches such as observations of independent evaluators, state-approved surveys of students and parents, or structured reviews of teacher artifacts of practice.				√ 40/60		

States	Comments	50/50	50/50 (roughly)	40/60 (roughly)	Equal across multiple categories	In development	No specific percentage
VA	State guidelines require seven performance standards. The first six, rated each at 10 percent, reflect the InTASC standards and the National Board for Professional Teaching Standards for practice; the seventh standard, student academic progress, accounts for 40 percent of the summative evaluation, of which at least 20 percent is comprised of student growth percentiles (SGPs), and another 20 percent using one or more alternative measures.			√ 40/10/ 10/10/ 10/10/10			
DE	Of the five component measures, student growth can be weighted only as high as the others.				√ 20/20/20/ 20/20		
KS	Building on work in progress to develop an evaluation system that is sensitive to the contextual challenges of Kansas educators (for example, isolated rural schools, hard-to-fill subject areas, and declining local school budgets). Guidelines are under development and a pilot is being conducted to determine artifacts that impact student achievement. Multiple measures examined include achievement on state assessments, observations, peer observations, professional growth, self-reflection, student and parent voice, and others.					√	
MO	The model system under development requires a minimum of three indicators: professional commitment, professional practice, and professional impact (includes measures of growth in student learning). Local education agencies are exploring ways through pilots to incorporate student growth in their local evaluation processes.					√	
SC	Currently the evaluation system, ADEPT, uses six measures of performance: teacher long-term plans; unit work samples to demonstrate student learning; classroom observations; teacher reflections following each classroom observation; professional performance review; and professional assessment, completed by the teacher as the first step to developing the teacher's professional growth and development plan. Additional performance measures (such as peer evaluations and student surveys) are being considered. The 59 schools in the Teacher Advancement Program (SC TAPTM) are serving as incubators for value-added assessments for teachers in tested subject areas and grades. Weighted values have yet to be determined.					√	

States	Comments	50/50	50/50 (roughly)	40/60 (roughly)	Equal across multiple categories	In development	No specific percentage
WA	The new law (ESSB 58975) sets forth eight evaluation criteria for teachers and requires student growth to be a “substantial factor” in a minimum of three of eight teacher criteria. The specific percentage to be attributed to student growth in the new evaluation systems has yet to be determined.					√	
AK	A certain percentage of student performance is not assigned to the overall evaluation in the state law, but it does specify that half of the evidence used must be student performance indicators that are externally generated or artifacts that the teacher has not designed or scored.						√
NC	Teacher contribution to student academic success is one of six evaluation standards. Currently there is no index or weighting system for the six standards. Failing to meet expectations on all six results in a status of “in need of improvement.”						√
OR	The evaluation framework includes three criteria: professional standards of practice, professional responsibilities, and student learning and growth. Guidelines for local systems are being developed.						√
RI	Evaluations must contain student growth, professional practice, and professional responsibility components. The growth model will not be used until there are two years of available assessment data. No required weights are established, though each local district’s evaluation system must base effectiveness “primarily” on evidence of student growth and academic achievement.						√

Source: The source is the state waiver applications referenced when we first mention the states on pp. 2-3.

The different categories of the components and their weight on the state evaluation systems can be described most simplistically as those states with:

- **50-50 percentage split between student-performance data and measures of professional practice used to determine overall effectiveness.** This categorization applies solidly to states such as Louisiana, which already has its guidelines in place, has completed pilots of various measures, is settled on a specific growth measure, and is ready for implementation for all teachers in the current school year. This contrasts with other approved waiver applicants that require extended qualification of the 50-50 commitment such as the District of Columbia’s Race

to the Top grantees or Michigan, which will meet the 50-50 split by increasing the percentage of the evaluation based on student growth in successive years.

- **Sliding percentages (a 50-50 percentage split roughly).** Arizona provides a sliding option of 33 percent to 50 percent tied to student quantitative data; an optional 17 percent tied to school-level and/or system-level data; and 50 percent to 67 percent reflective of professional practice.
- **40/60 percentage split (roughly).** Utah, with growth measures under development, has set a floor of 40 percent for student growth. The remaining 60 percent specifies 40 percent for professional practice based on observations and a likely 20 percent for parent/student input. Connecticut specifies 45 percent of the composite evaluation be based on standardized assessments of student performance. On the professional practice side, Connecticut specifies 40 percent be based on teacher observation and/or professional practice, 10 percent on peer and parent feedback, and 5 percent on school-wide student-learning indicators or student feedback, all of which totals 55 percent of the composite for measures of professional practice.
- **Equal weights across categories.** In Delaware, student growth can be weighted only as high as the four other component measures.
- **Assignment of weights under development.** This is the case for Kansas, Missouri, South Carolina, and Washington.
- **No specific assigned weights.** In Arkansas, North Carolina, Oregon, and Rhode Island, there is either no assigned weighting of evaluation components in state law, no mention of specific weights, or none have been established.

Making significant progress

The state actions discussed in this report reflect a period in the efforts of states to build systems of teacher evaluation and support. Of note is how far many of the states have come in their approach in contrast to the teacher evaluation landscape of only a few years ago. By their own admissions, states have moved beyond checklists of teacher performance, and showcase-lessons where a “pass” or “satisfactory” was the given and evaluation consequently did little to improve instructional practices. States are building consistent and uniform standards of quality into their systems where few existed before. Many are taking novel approaches to old ideas, including building systems that are more closely attuned to the career development of educators. Most are clearly linking professional development components to evaluations in ways that have not been done in the past and are planning how to finance professional development. Furthermore, many states acknowledge improvements in student achievement as the driving goal of their evaluation systems, and one in particular—Rhode Island—has made a strong commitment to making sure every student has access to an effective teacher.

The roles played by the states are influenced by such factors as the characteristics of local school districts, the laws governing charter school autonomy, the balance between local control and state autonomy, and collective bargaining agreements related to educator evaluation. As a result, states have had to gauge the components of their evaluation systems that should be mandated and those that should be left to the discretion of local school districts, while at the same time maintaining the integrity of a comprehensive and consistent statewide approach. This decision ultimately shapes the roles and responsibilities of states and the capacity required to do the work.

Let’s examine specific examples of how this change has manifested itself in several states.

Standardizing evaluation practices statewide

- **Arkansas:** Local districts in the state previously chose or designed their own teacher and administrator evaluation instruments. There were consequently no consistent standards or a uniform system for the support and improvement of teacher effectiveness. Arkansas has now developed a standard evaluation process that honors local flexibility to adopt, adapt, or modify the standard evaluation to meet local needs that are consistent with the state model. The state now describes its new evaluation system as a “significant part of a comprehensive and coherent differentiated system for accountability, recognition and tiered support.”⁶⁶
- **Washington:** Educators in this state have received annual evaluations for more than 30 years. Evaluation systems were developed and bargained locally and were completed at the discretion of each district. Though the new state law has yet to be fleshed out and the state is still in a start-up phase, the law creates one state model with specific and consistent choices for districts to consider as they construct their teacher evaluation systems.
- **Connecticut:** The state’s Senate Bill 458 requires different professional development activities based on evaluation results and diverges from previous law where professional development was based largely on “seat-time” or continuing education units. Districts are required to provide job-embedded, effective professional development that focuses on strengths and needs identified through the model evaluation system, but they have the flexibility to design customized professional development based on evaluation data and focused on individual teacher needs. Districts are in turn held accountable for providing professional development that effectively meets the needs of educators, especially those with the greatest need for support. Connecticut is among states that have made effectiveness a requirement for tenure.
- **Louisiana:** The state’s new teacher and leader evaluation and support system radically differs from earlier systems that only measured teacher competencies in the classroom. The new system ties educator performance to student achievement, allows educators to set meaningful and ambitious professional and student achievement goals, and supports a comprehensive system of observation, evaluation, and feedback to guide professional development that is specific to teacher needs and goals.

- **Maryland:** The state’s theory of action—the underlying assumptions about how it will move from the present to a stronger and more effective education system—uses professional development as the foundation for improving and maintaining educator effectiveness. Maryland urges its local school districts to use federal Elementary and Secondary Education Act Title II, Part A funds for professional development, as well as local funds to support professional learning that is directly aligned with the qualitative components of the teacher evaluation system.
- **North Carolina:** North Carolina’s theory of action is that every student should have effective teachers and that every school should have an effective leader. Definitions of effective and highly effective teachers and leaders have been established and will be infused into new policies governing a range of important areas, including career status or tenure, licensing, retention and dismissal, incentives for equitable teacher and leader distribution, and evaluation of teacher and leader preparation programs.
- **Ohio:** The state’s House Bill 153, which established a standards-based state framework for the evaluation of teachers, also provides funding for professional development to support teacher growth and the development of poorly performing teachers.
- **Rhode Island:** The state has indicated that every human resource decision made in regard to educators in the state—whether by a local education agency or the state education agency—will be based on evidence of the respective teacher’s or principal’s impact on student growth and academic achievement, as well as other measures of professional practice and responsibility.⁶⁷
- **South Carolina:** The state’s evaluation system, originally adopted in 2006, has been refined to comply with the Elementary and Secondary Education Act flexibility request. Prior to this system, evaluation instruments were for the most part limited to behavioral checklists and showcase lessons. Almost all teachers passed these evaluations, and the evaluation did little to improve instructional practices. The current system, in contrast, is designed as an iterative process rather than a final product. Its performance standards define the expectations for teacher effectiveness through the entirety of a teacher’s career. The standards apply to the preparation of teacher candidates, as well as each stage of teacher practice. These stages encompass induction and mentoring for first-year teachers and formal evaluation for certification, contract advancement, high-stakes personnel decisions, and goals-based evaluation for experienced educators. The

South Carolina Department of Education created a new office—for educator evaluation—to demonstrate the high priority of this work. It is also invested in the informational and reporting needs of the system and has created web-based systems on the annual performance of every teacher and principal in the state. The data system enables districts to compare the performance of their teachers at each contract level with the performance of teachers statewide.

Challenges still ahead

A number of states have clearly articulated belief systems that link the quality of educator evaluation systems to the quality of student learning, define educator effectiveness, and recognize the importance of meaningful feedback and targeted professional development for educator improvements. Having these beliefs and theories of action codified into state law, regulations, or other standards of practice indicates the strength of the game change we are now witnessing. There are, however, still challenges ahead—challenges related to the selection and application of tools, the chosen design elements, the training of evaluators to ensure consistency, the common understanding and the necessary buy-in of stakeholders about the purposes and applications of the measures selected, and the implementation and continual refinements of the systems.

Additionally, there are a number of technical challenges that must be addressed: defining and measuring teaching behavior; gathering information through consistent and reliable observation; ensuring that the teacher behaviors observed really matter for student learning; determining how observations connect to high-stakes consequences such as tenure and professional development; and a host of support and infrastructure requirements needed to roll-out sound observation efforts on a large scale.⁶⁸ There are also the concerns around value-added models that were discussed previously. And, of course, as states add other measures of student success, including student-learning objectives, the rigor and consistency of these measures must be ensured.

For several states, much of the major design work is still under development. This includes the work of the districts to tailor their evaluation plans to state requirements and the work of the states to ensure that district systems fall within these requirements. We can already witness the magnitude of the challenge ahead. Each of New York's roughly 700 school districts must have a state-approved teacher and principal evaluation plan in place by January 17, 2013, yet many districts are still negotiating

details with their teachers unions.⁶⁹ As of September 19, 2012, the state had only approved 107 district plans and a logjam is foreseeable as the deadline nears.⁷¹

As they make their way toward reform, the second round of waiver applicants have been able to benefit from the work of the earlier implementers. One of the second-round states, Utah, has been working with Colorado and other early adopters of student-growth measures (Delaware, Georgia, and Rhode Island) before phasing in its growth measures in the 2013–14 school year. Nevada’s growth model is based on the Colorado growth model. And a number of other states are looking to the experiences of their School Improvement Grant and Teacher Incentive Fund grantees for important lessons to apply to their new teacher evaluation and support systems.

Understandably, the road to reform has potholes and some districts and states have run into controversies. This became clear during the recent teacher’s union strike in Chicago, where the use of student test scores in teacher evaluations, as stipulated by the state, surfaced as one point of contention. This factor, however, seems to have settled into an accepted component of the teacher contract in line with state law. Where there is contention, concerns often relate to whether these new teacher evaluation systems are fair, reliable, and valid. Allaying these concerns is often critical to the success of these new systems. The earlier implementer states have potential to offer lessons on this subject. The willingness to move ahead, but only after stepping back, taking stock, and recalibrating, will likely give further direction to these new efforts.⁷¹

Throughout the reform process, one issue has gone largely unaddressed: How states would tackle the Elementary and Secondary Education Act requirement (one not exempted by the waiver) that poor and minority children not be taught by unqualified, inexperienced, or out-of-field teachers at higher rates than other children.⁷² The hope is that these improved systems of educator evaluation and support may become the tools to rectify these inequities by improving the overall quality of the field. Rhode Island has committed to the goal that no child in the state “will be taught by a teacher who has been rated ineffective for two consecutive years.”⁷³ This commitment bears watching, as do many of the aspirational claims mentioned by other states, once the spotlight is removed and other policy priorities take center stage. It is our hope that in time, all states can and will commit to the goal that all children should be taught by an effective teacher every year. This outcome, however, has yet to be realized.

Findings and recommendations

A review of these various state plans indicates that the design and implementation of new systems of evaluation and support is truly a work in progress. It is clearly hard work to legislate, regulate, and provide guidance for change within an environment of multiple simultaneous reforms. These reforms include the implementation of new college- and career-ready standards, statewide data systems, new assessments, and new state responsibilities for these new systems, to name a few—all challenging an established status quo and each bearing on the other.

It is evident from reading these plans that each state approach is different and that each is in a different place in terms of development and implementation, although some of the second round states have benefited from the work of the early implementers and most are benefitting from the modeling and pilots already in place prior to full statewide implementation. Teacher evaluation designs are influenced by factors such as the characteristics of local school districts, laws governing charter school autonomy, and the state history for local control and collective bargaining agreements related to educator evaluation.

It's also clear that states are relying on a range of measures and methods for assessing teacher professional practice. These include classroom observations, self-assessments and reflection, teaching artifacts, student-learning measures, and surveys of students and parents. In addition to measures of professional practice, waiver winners are using both student achievement and growth measures, including value-added estimates when available, to capture measures of student success aligned with individual teachers or teams of teachers. A number of states are still considering the types of student-growth measures to use, and some are piloting multiple models before they recommend a particular approach.

In fine-tuning their approaches, states are also looking to more personalized and school-appropriate measures for determining teacher impact on student learning and vesting teachers more directly in monitoring student progress. Whether called student-achievement goal setting, student-learning objectives, student-learning

targets, teacher goal setting, or unit work samples, these measures are used to actively engage the teacher and the evaluator in a goal-setting process for student learning that is customized for the teaching assignment and for the students.

The review of state plans shows that states use different measures and give different weights to measures for student achievement and professional practice. Some states have specific percentages of components spelled out in state law; others do not. In some cases a certain amount of discretion is given to local districts for insertion of components they value in the evaluation.

All states are determining or developing assessments applicable to teachers of grades and subjects that are not part of statewide standardized assessments for the purpose of determining student growth. They are expanding the portfolio of state assessments to provide growth data in all grades and subjects or expanding the portfolio of nationally or locally approved assessment tools that can be validly used such as classroom-based assessments, unit tests, end-of-course assessments, student-learning objectives, and portfolios.

It's heartening to see that waiver applicants were responsive to the application requirements to make these systems as much about differentiating educators on their levels of effectiveness—for use in making personnel decisions—as about letting the evaluation process be a larger part of a system of supports for overall improvement. Many states had already started along this pathway when the waiver requirements represented an opportunity to tweak existing designs. For others, the Race to the Top and waiver requirements represented an opportunity to insert measures of student success into the components of the overall evaluation and to create an aligned role for professional development and peer assistance. Many states have progressed a long way in a relatively short period of time, and are now building consistent and uniform standards of quality where few existed before.

At the end of the day, it is not just about building new systems of teacher evaluation but also is about ensuring that the infrastructure is in place to ensure the success of these systems. This means that teachers and principals must receive orientation to the new systems; evaluators must receive appropriate training (for example, in collecting evidence-rating against a professional standard and providing feedback); rubrics and protocols for high-quality observations must be identified and tested; strong student-teacher data links, evaluation reporting systems, and quality controls for verifying the accuracy and reliability of evaluations must be in place; and management systems must be devised that allow teachers to track

their progress toward learning goals. Just as importantly, supports and interventions must be in place to move teachers toward higher levels of effectiveness in line with the information provided through evaluation.

This entire process will likely be an iterative one and should be open to review and adjustment as new research and the results of pilot implementation surface. Certainly, no one expects the design and implementation of these systems to be perfect in their first or second attempts. For now, the state efforts and the waiver process represent a rich laboratory of exploration and reform that merit watching—both for lessons to be learned and also for necessary corrections to be made.

The next iteration of these systems of support will likely be much more refined, as the field assigns specific measures to the realms where they provide the best information: as a basis of personnel action; to support teacher professional growth and development; as a mechanism for aligning teacher and student effort and goals; and as a way of leveraging educator strengths and allowing for differentiated job responsibilities. These systems may move from the search for one composite number or score representing all these purposes to a more complex structure of triggers, indicators, and aligned interventions.

Against this evolving backdrop we offer the following recommendations:

- The U.S. Department of Education should closely monitor the successes and problems experienced by these states and the District of Columbia as they implement these new systems of teacher evaluation and support. Some of the approved applications lacked detail, and many components and decisions were still in the developmental phases. As states finalize or change their evaluation policies, the department must ensure states comply with waiver requirements and maintain rigorous standards. Some states may be tempted to take a less resistant path, especially since they have already received a waiver, and this must be closely monitored.
- The states and the District of Columbia should continue to heed emerging findings from research and evaluation and seek feedback from their own efforts to ensure continuous improvements. The department can help by creating a clearinghouse of best practice and perhaps communities of practice in the way it has done for the Race to the Top grantees.

- The U.S. Department of Education and philanthropic organizations should continue to support improvements in the tools and infrastructure necessary for the development and sustainability of these new systems. Existing funding streams such as federal Title II A of the Elementary and Secondary Education Act and local funds for professional development should be reviewed for how they can support this work.
- Lessons learned from these efforts should provide critical information on the needs and capabilities of the states and districts to improve and support future direction for the reauthorization of the Elementary and Secondary Education Act.

Conclusion

These are exciting times, as a number of federal, state, and local initiatives are in a position to change the face of education. Key to this transformation is the important work being done with respect to teacher evaluation—work that, if done well and embraced by educators, will provide the foundation for strong human capital management systems that will help build strong faculties and schools capable of supporting student learning in our nation’s public schools.

Undeniably, this is hard work requiring dedication from and commitment of all the stakeholders, but the potential payoffs are huge. At this point, the outcomes of the waiver process and of the policies being pursued by the states and the federal government in relation to systems of evaluation and support remain to be seen. It is not a stretch to say, however, that the successful reform of teacher evaluation will finally give teachers the support and feedback they need to be successful; give school leaders the fact-based data they need to make informed personnel decisions; and, most importantly, give students the effective teachers they need to achieve academically.

About the author

Glenda L. Partee is the Associate Director for Teacher Quality at the Center for American Progress. Her work focuses on improvements in human capital systems in our public schools. Prior to joining American Progress, she was an independent education consultant who advised and wrote for local and state school systems, education associations, foundations, and nonprofit organizations on diverse issues. From 2005 to 2009, Partee served in a number of capacities at the District of Columbia Office of the State Superintendent of Education, including as director of policy, research, and analysis, and assistant superintendent for postsecondary education and workforce readiness. Previously, she was co-director of the American Youth Policy Forum and held positions at the Council of Chief State School Officers and the National Association for Equal Opportunity in Higher Education. She was a member of the New York City Urban Teacher Corps and taught in schools in New York City and St. Croix, U.S. Virgin Islands. Partee has a doctorate in instructional systems from Pennsylvania State University, a master's degree from the City College of New York, and a bachelor's degree from Mt. Holyoke College.

Endnotes

- 1 Daniel Weisberg and others, "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness" (Brooklyn, New York: The New Teacher Project, 2009). The report discussed the ways in which teachers evaluation systems reflect and reinforce indifference to variations in teacher performance wherein all teachers are rated good or great; excellence goes unrecognized; professional development is inadequate; no special attention is given to novices; and poor performance goes unaddressed.
- 2 The Education Trust-West, "Learning Denied: The Case for Equitable Access to Effective Teaching in California's Largest School District" (2012).
- 3 According to Robert C. Pianta, dean of the Curry School of Education at the University of Virginia, "Relying on the status quo for teacher performance evaluation wastes time and energy—performance metrics are nonexistent or not valid and there is little to no linkage among the key components of most evaluation and performance-improvement systems. As practiced now teacher evaluation is a nonsystem with a lot of moving parts of dubious value and very little connection among them." Robert C. Pianta, "Implementing Observation Protocols: Lessons for K-12 Education from the Field of Early Childhood" (Washington: Center for American Progress, 2012).
- 4 Colorado, Florida, Georgia, Indiana, Kentucky, Massachusetts, Minnesota, New Jersey, New Mexico, Oklahoma, and Tennessee were in the first round of applicants that received flexibility from No Child Left Behind.
- 5 Department of Education, *ESEA Flexibility Request: Arizona Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/az.pdf>.
- 6 Department of Education, *ESEA Flexibility Request: Arkansas Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ar.pdf>.
- 7 Department of Education, *ESEA Flexibility Request: Connecticut Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ct.pdf>.
- 8 Department of Education, *ESEA Flexibility Request: Delaware Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/de.pdf>.
- 9 Department of Education, *ESEA Flexibility Request: District of Columbia Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/dc.pdf>.
- 10 Department of Education, *ESEA Flexibility Request: Kansas Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ks.pdf>.
- 11 Department of Education, *ESEA Flexibility Request: Louisiana Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/la.pdf>.
- 12 Department of Education, *ESEA Flexibility Request: Maryland Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/md.pdf>.
- 13 Department of Education, *ESEA Flexibility Request: Michigan Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/mi.pdf>.
- 14 Department of Education, *ESEA Flexibility Request: Mississippi Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ms.pdf>.
- 15 Department of Education, *ESEA Flexibility Request: Missouri Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/mo.pdf>.
- 16 Department of Education, *ESEA Flexibility Request: Nevada Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/nv.pdf>.
- 17 Department of Education, *ESEA Flexibility Request: New York State Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ny.pdf>.
- 18 Department of Education, *ESEA Flexibility Request: North Carolina Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/nc.pdf>.
- 19 U.S. Department of Education, *Ohio ESEA Flexibility Request* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/oh.pdf>.
- 20 Department of Education, *ESEA Flexibility Request: Oregon Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/or.pdf>.
- 21 Department of Education, *ESEA Flexibility Request: Rhode Island Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ri.pdf>.
- 22 Department of Education, *ESEA Flexibility Request: South Carolina Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/sc.pdf>.
- 23 Department of Education, *ESEA Flexibility Request: South Dakota Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/sd.pdf>.
- 24 Department of Education, *ESEA Flexibility Request: Utah Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/ut.pdf>.
- 25 Department of Education, *ESEA Flexibility Request: Virginia Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/va.pdf>.
- 26 Department of Education, *ESEA Flexibility Request: Washington State Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/wa.pdf>.
- 27 Department of Education, *ESEA Flexibility Request: Wisconsin Department of Education* (2012), available at <http://www2.ed.gov/policy/eseaflex/approved-requests/wi.pdf>.

- 28 No Child Left Behind requires that all public school teachers of core academic subjects meet the “highly qualified teacher” requirements of the act. To be highly qualified, a teacher must hold a bachelor’s degree, have obtained full state certification, and have demonstrated subject-matter expertise in each core academic subject taught. States are required to report to the U.S. Department of Education on progress toward the 100 percent HQT goal.
- 29 American Institutes for Research, “Reauthorizing ESEA: Making Research Relevant” (2011).
- 30 Primarily as a result of this competition, 33 states have recently passed teacher evaluation legislation—each with the goal of improving the quality of instruction in schools. National Council on Teacher Quality, “State of the States: Trends and Early Lessons on Teacher Evaluation and Effectiveness Policies” (2011), available at http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf.
- 31 For example, Virginia’s percentage of highly qualified teachers was 99.3 percent for the 2010–11 school year; for its high poverty schools, the percentage was 98.8 percent. Department of Education, *ESEA Flexibility Request: Virginia Department of Education*, p. 71.
- 32 Department of Education, *ESEA Flexibility Request: Maryland Department of Education*, p. 78.
- 33 See, for example, the findings from Year 2 of the Measures of Effective Teaching project to test multiple measures of teacher effectiveness. The project analyzes five measures of effectiveness to help establish which combination captures the full range of teacher contributions to student learning. The research components include student achievement gains on state standardized assessments, as well as supplemental assessments, to measure higher-order conceptual thinking, classroom observations and teacher reflections, teachers’ pedagogical content knowledge, student perceptions of the classroom instructional environment (measured through student surveys), and teachers’ perceptions of working conditions and instructional support at their schools. Measures of Effective Teaching Project, “A Composite Measure of Teacher Effectiveness” (2010).
- 34 Department of Education, *ESEA Flexibility FAQ* (2012), p. 31–32.
- 35 Department of Education, *ESEA Flexibility Review Guidance* (2012), p. 19.
- 36 Laura Goe, Kietha Biggers, and Andrew Croft, “Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning” (Washington: National Comprehensive Center for Teacher Quality, 2012); Kelly Burling, “Evaluating Teachers and Principals: Developing Fair, Valid, and Reliable Systems” (Hoboken, New Jersey: Pearson Education, Inc., 2012).
- 37 The “*InTASC Model Core Teaching Standards: A Resource for State Dialogue*” outlines what all teachers across all content and grade levels should know and be able to do to be effective in today’s learning contexts. They are a revision of the 1992 model standards and describe a new vision of teaching designed to meet the needs of the next of generation learners. Council of Chief State School Officers, “*InTASC Model Core Teaching Standards: A Resource for State Dialogue*” (2011).
- 38 “The Framework for Teaching” is often used as the foundation for dialogue among practitioners and in the mentoring, coaching, professional-development, and teacher-evaluation processes. The four domains addressed in the framework are planning and preparation, classroom environment, instruction, and professional responsibilities. States may use the Danielson Framework to provide definition and specificity to the InTASC standards. “The Framework for Teaching,” available at <http://www.danielsongroup.org/article.aspx?page=frameworkforteaching> (last accessed October 2012).
- 39 According to Robert C. Pianta, there is little data indicating the appropriateness of cut-off scores separating “sufficient” from “insufficient” levels of teaching skill. There are also no published norms to guide expected levels of change in response to interventions over time. It is therefore important to be cautious in using observational data to determine whether teachers pass or fail in the quality of their teaching or whether their progress in response to intervention is sufficient or lacking. Pianta, “Implementing Observation Protocols.”
- 40 Burling, “Evaluating Teachers and Principals”; The New Teacher Project, “Teacher Evaluation 2.0” (2010).
- 41 Pianta, “Implementing Observation Protocols.”
- 42 Measures of Effective Teaching Project, “A Composite Measure of Teacher Effectiveness” (2010).
- 43 The study involved nearly 3,000 teacher-volunteers evaluating alternative ways to provide valid and reliable feedback to teachers for professional development and improvement. The five instruments invested in the study were framework for teaching, classroom assessment scoring system, protocol for language arts teaching observations, mathematical quality of Instruction, and UTeach teacher observation protocol. Ibid.
- 44 Measures of Effective Teaching Project, “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains” (2012).
- 45 “Peer Review of Teaching,” available at <http://www1.umn.edu/ohr/teachlearn/resources/peer/index.html> (last accessed October 2012).
- 46 Measures of Effective Teaching Project, “A Composite Measure of Teacher Effectiveness.”
- 47 As described on the District of Columbia Public Schools website, expert master educators are “talented leaders with a proven track record of success in making schools work for students, families and communities.” They “serve as impartial, third-party evaluators of teacher performance”; “provide teachers with targeted, content-specific feedback and resources”; and “provide instructional capacity to support DCPS reform initiatives.” See: “Master Educators,” available at <http://dcps.dc.gov/DCPS/About+DCPS/Career+Opportunities/Lead+Our+Schools/Master+Educators>.
- 48 Virginia, for example, allows for the use of Student Achievement Goal Setting as a measure of student growth when valid measures of student academic progress are not available. Goal setting is used to focus attention on students and on instruction by determining baseline performance, developing strategies for improvement, and assessing results at the end of the academic year. Student academic progress goals are used to measure where the students are at the beginning of the year, where they are at mid-year, where they are at the end of the year, and the difference between all three. Appropriate measures of student-learning gains may include criterion-referenced tests, norm-referenced tests, standardized achievement tests, school-adopted interim, common, or benchmark assessments, authentic measures (e.g., learning portfolio, recitation, and performance), and teacher-generated measures of student performance (e.g., teacher developed assessments and performance-based as-

- sessments). Department of Education, *ESEA Flexibility Request: Virginia Department of Education*, p. 120–121.
- 49 Kimberly O'Malley and others, "Overview of Student Growth Models" (Hoboken, New Jersey: Pearson Education, Inc., 2011).
- 50 Kimberly O'Malley and others, "Making Sense of the Metrics: Student Growth, Value-Added Models, and Teacher Effectiveness," *Pearson Assessments Bulletin* (19) (2011).
- 51 Jennifer L. Steele, Laura S. Hamilton, and Brian M. Stecher, "Incorporating Student Performance Measures into Teacher Evaluation Systems" (Santa Monica and Washington: RAND Corporation and the Center for American Progress, 2010).
- 52 These reasons were among those cited by Georgia professors in an open letter to Governor Nathan Deal. Valerie Strauss, "Georgia professors blast teacher evaluation system," *The Washington Post*, July 10, 2012, available at http://www.washingtonpost.com/blogs/answer-sheet/post/georgia-professors-blast-teacher-evaluation-system/2012/07/09/gQAfh5bZW_blog.html; Linda Darling-Hammond, "Value-Added Evaluation Hurts Teaching," *Education Week*, March 5, 2012, available at http://www.edweek.org/ew/articles/2012/03/05/24darlinghammond_ep.h31.html.
- 53 Research is beginning to show the relationship among multiple measures and the relative strengths of different measures. A Consortium on Chicago School Research study, for example, found a strong relationship between classroom observation ratings and value-added measures, with students in the classrooms of highly rated teachers showing the most growth, and students in classrooms of teachers with low observation ratings showing the least growth. Lauren Sartain and others, "Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation" (Chicago: Consortium on Chicago School Research at the University of Chicago Urban Education Institute, 2011); Measures of Effective Teaching Project, "A Composite Measure of Teacher Effectiveness"; Diana Epstein and Raegen Miller, "Subtraction by Distraction: Publishing Value-Added Estimates of Teachers by Name Hinders Education Reform" (Washington: Center for American Progress, 2011).
- 54 Douglas N. Harris, "How Do Value-Added Indicators Compare to Other Measures of Teacher Effectiveness?" *Carnegie Knowledge Network*, October 15, 2012, available at http://www.carnegieknowledge-network.org/briefs/value-added/value-added-other-measures/?utm_source=CKN+Mailing+List&utm_campaign=3588d32a7e-CKN_announcement10_19_2012&utm_medium=email.
- 55 The Partnership of Assessment of Readiness for College and Careers is a 23-state consortium working together to develop the next generation of K–12 assessments in English and math.
- 56 Cohort 1 includes English language arts, mathematics, science, social studies, and world languages. Cohort 2 includes English as a Second Language, health, physical education, music, and visual and performing arts. Cohort 3 includes family and consumer science, business, finance and marketing, technology education, health sciences, agriculture, and skilled and technical sciences. Cohort 4 includes the following nonsubject educators: counselors, librarians, physical and occupational therapists, educational diagnosticians, speech pathologists, psychologists, nurses, visiting teachers, and preschool and special education teachers involved in alternative assessments. Department of Education, *ESEA Flexibility Request: Delaware Department of Education*, p. 117.
- 57 The Index includes student achievement data in English language arts, math, and science for prekindergarten through grade 12, and growth data in English language arts and math for prekindergarten through grade eight. For grades nine through 12, the index includes high school graduation and dropout rates. Department of Education, *ESEA Flexibility Request: Maryland Department of Education*, p. 76.
- 58 Department of Education, *ESEA Flexibility Request: Maryland Department of Education*, p. 180. Metrics that serve as the basis of the evaluation for student growth are based on courses and grade levels as follows:
- For elementary and middle school teachers who teach more than one subject, student growth would be calculated by combining the aggregate of 10 percent of the class reading scores on the on Maryland State Assessment, 10 percent of the class math state assessment, 20 percent of the student-learning objectives, and 10 percent from the School Performance Index.
- For elementary and middle school teachers who only teach one subject student, growth is calculated using 20 percent from student-learning objectives, 10 percent from the School Performance Index, and the final 20 percent from the class scores of the appropriate subject.
- For elementary or middle school teachers who teach in nontested content area, student growth is determined by the student-learning objectives (35 percent) and the School Performance Index rating (15 percent). These same multiple measures are used for high school teachers.
- 59 Districts may locally bargain the selection of these measures and the process for assigning points to educators. Allowable options include measures based on state assessments, Regents examinations and/or state-approved alternatives to Regents examinations (provided that the measures are different from the measures used for the growth subcomponent), measures based on the state-approved list of third-party assessments, measures based on district, regional, or Board of Cooperative Educational Services development assessments, school-wide growth or achievement results, or student-learning objectives. Department of Education, *ESEA Flexibility Request: New York Department of Education*, p. 139.
- 60 The SAS Education Value-Added Assessment model for K–12 uses a longitudinal analysis to track individual student progress by year, grade, and subject based on a variety of assessments. "SAS EVAAS for K-12," available at <http://www.sas.com/govedu/edu/k12/evaas/index.html> (last accessed November 2012).
- 61 Public Schools of North Carolina, "Measuring Growth for Educator Effectiveness" (2012), available at <http://www.ncpublicschools.org/docs/educatoreffect/ncees/measure-growth-guide.pdf>.
- 62 These include end-of-grade and end-of-course exams in grades three through eight in English language arts; one year of high school English language arts; grades three through eight in mathematics, one year of high school mathematics; grades five and eight in science; high school biology; and summative post-assessments for all career and technical education courses. Department of Education, *ESEA Flexibility Request: North Carolina Department of Education*, p. 121.
- 63 Personal communication from Jennifer Preston, Race to the Top Project Coordinator for Teacher and Leader Effectiveness, North Carolina Department of Public

- Instruction, December 9, 2012. Since the original waiver application was submitted, the North Carolina Department of Public Instruction has decided to use the roster verification tool in the Education Value-Added Assessment System, or EVAAS, web interface. All North Carolina teachers have EVAAS accounts and can access the system to verify their class lists. In doing so, teachers can also indicate when they share responsibility for the instruction of one or more students. The EVAAS model is then able to weight the growth of those shared students appropriately and include it in the teacher's individual value-added score, which takes away the need to have a team value-added score.
- 64 The unit work sample is based on the teacher work-sample concept developed by Renaissance Partnership for Improving Teacher Quality. In addition to determination of major unit objectives (a unit is defined as a set of integrated lessons designed to accomplish learning objectives related to one or more curricular themes, areas of knowledge, and/or general skills or processes) and an instructional plan, unit assessments (formative and summative) reflect student achievement growth. Department of Education, *ESEA Flexibility Request: Rhode Island Department of Education*, p. 148.
- 65 Matthew Di Carlo, "Teacher Evaluations: Don't begin assembly until you have all the parts," Shanker Blog, July 19, 2011, available at <http://shankerblog.org/?p=3165>.
- 66 Department of Education, *ESEA Flexibility Request: Arkansas Department of Education*, p. 141.
- 67 Inclusive of certification, selection, tenure, professional development, support for individual and groups of educators, placement, promotion, compensation, and retention. Department of Education, *ESEA Flexibility Request: Rhode Island Department of Education*.
- 68 Pianta, "Implementing Observation Protocols."
- 69 New York is a local-control state, and districts must collectively bargain many aspects of their evaluation systems. The state balances a need for local flexibility with the use of consistent design elements associated with improved student learning and teacher practice. For these reasons, the state's role is focused on developing statewide measures of student growth, determining how growth will be measured in subjects where there are no state assessments, approving locally selected third-party assessments, rubrics of educator practice, and student and parent survey tools, delivering training and resources for turn-key local training, and providing guidance and support to districts as they plan their systems and meet the requirements of the law. (N.Y. Educ. § 3012-c); Department of Education, *ESEA Flexibility Request: New York State Department of Education*, p. 139–140.
- 70 "NY Has OK'd 107 Teacher Evaluation Plans," *Times Herald-Record*, September 20, 2012, available at <http://www.recordonline.com/apps/pbcs.dll/article?AID=/20120920/NEWS90/120929997/-1/rss01>; "N.Y. Districts Approach Third Deadline for Teacher-Evaluation Plans," available at http://article.wn.com/view/2012/09/18/NY_Districts_Approach_Third_Deadline_for_TeacherEvaluation_P/.
- 71 For example, Tennessee, a Race to the Top state and first-round waiver state, noted varied satisfaction with its new evaluation system among districts during early implementation and public discussion that began to detract from the purpose of the evaluation system: to improve student achievement. To address these concerns, the state undertook an extensive statewide listening and feedback process that has resulted in major recommendations affecting the design and implementation of the system. These recommendations include an examination of the components of the 50 percent of the evaluation scores driven by student achievement data (currently 35 percent is based on student growth on the state test or comparable measure, and 15 percent is based on additional measures of student achievement data adopted by the State Board of Education and chosen by the mutual agreement of the educator and evaluator); changes to the qualitative rubric to improve discussion and feedback about improvements in instruction; increases in process efficiencies so that administrator time is better spent on observations and teacher feedback; and other quality control approaches effecting evaluators. Tennessee Department of Education, "Teacher Evaluation in Tennessee: A Report on Year 1 Implementation" (2012).
- 72 Department of Education, *ESEA Flexibility: Frequently Asked Questions* (2012), p. 27, available at <http://www2.ed.gov/policy/eseaflex/esea-flexibility-faqs.doc>.
- 73 Department of Education, *ESEA Flexibility Request: Rhode Island Department of Education*, p. 113.

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just, and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is “of the people, by the people, and for the people.”

Center for American Progress

