



Evaluating Teacher Effectiveness

How Teacher Performance Assessments Can Measure and Improve Teaching

Linda Darling-Hammond October 2010

Center for American Progress



Evaluating Teacher Effectiveness

How Teacher Performance Assessments Can Measure and Improve Teaching

Linda Darling-Hammond October 2010

Contents

1 Introduction and Summary

5 Why performance assessments are needed

- 7 National Board Certification
- 9 Assessment of beginning teachers

11 Prospects for a national system of Teacher Performance Assessments

- 12 How performance assessments can help teachers improve their practice
- 17 Improving preparation through the use of performance assessments

22 Policy recommendations and conclusion

Introduction and Summary

Parents, practitioners, and policymakers agree that the key to improving public education in America is placing highly skilled and effective teachers in all classrooms. Yet the nation still lacks a practical set of standards and assessments that can guarantee that teachers, particularly new teachers, are well prepared and ready to teach.

This report discusses a promising approach to the question of how to measure teacher effectiveness. Specifically, it describes the ways in which assessments of teacher performance for licensing and certification can both reflect and predict teachers' success with children so that they can not only inform personnel decisions, but also leverage improvements in preparation, mentoring, and professional development. It outlines progress in the field of teacher assessment development and discusses policies that could create much greater leverage on the quality of teacher preparation and teaching than has previously existed in the United States.

For more than two decades, policymakers have undertaken many and varied reforms to improve schools, ranging from new standards and tests to redesigned schools, new curricula and new governance models. One important lesson from these efforts is the repeated finding that teachers are the fulcrum determining whether any school initiative tips toward success or failure. Every aspect of school reform depends on highly skilled teachers for its success. This is especially true as educational standards rise and the diversity of the student body increases. Teachers need even more sophisticated abilities to teach more complex curriculum to the growing number of public school students who have fewer educational resources at home, those who are new English language learners, and those who have distinctive learning needs.

One of the few areas of consensus among education policymakers, practitioners, and the general public today is that improving teacher quality is one of the most direct and promising strategies for improving public education outcomes in the

United States, especially for groups of children who have historically been taught by the least qualified teachers. Teachers can have large effects on student achievement, as suggested by a recent large-scale study in North Carolina, which found that the differences in achievement gains for students who had the most qualified teachers versus those who had the least qualified were greater than the influences of race and parent education combined.¹ These very large differences were associated with teachers' initial preparation for teaching, licensing in the field taught, strength of academic background, level of experience, and demonstration of skills through National Board Certification, all of which are variables that could be directly addressed through policy.

Unlike most high-achieving nations, however, the United States has not yet developed a national system of supports and incentives to ensure that all teachers are well prepared and ready to teach all students effectively when they enter the profession.² Nor is there a set of widely available methods to support the evaluation and ongoing development of teacher effectiveness throughout the career, along with decisions about entry and continuation in the profession. Meeting the expectation that all students will learn to high standards will require a transformation in the ways in which our education system attracts, prepares, supports, and develops expert teachers who can teach in more powerful ways—a transformation that depends in part on the ways in which these abilities are understood and assessed.

In recent years, there has been growing interest in moving beyond traditional measures of teacher qualifications, such as completion of a preparation program, number of degrees, or years of experience, in order to evaluate teachers' actual performance as the basis for making decisions about hiring, tenure, licensing, compensation, and selection for leadership roles. A key problem is that current measures for evaluating teachers are not often linked to their capacity to teach. Existing federal, state, and local policies for defining and measuring teacher quality either rely almost exclusively on classroom observations by principals who differentiate little among teachers and offer little useful feedback, or focus on teachers' course-taking records and on paper-and-pencil tests of basic academic skills and subject matter knowledge that are poor predictors of later effectiveness in the classroom.

Looking ahead to the reauthorization of the Elementary and Secondary Education Act, the No Child Left Behind Commission called for moving beyond the designation of teachers as “highly qualified” to an assessment of teachers as “highly effective” based on student learning evidence. Other recent federal proposals—for example, the Teacher Excellence for All Children Act—have suggested incentive

pay to attract effective teachers to high-need schools and to pay them additional stipends to serve as mentors or master teachers. The questions are now squarely on the table: “How should we measure teacher effectiveness?” And how can we develop more effective teachers much more consistently, rather than leaving teacher effectiveness to chance?

This report describes progress currently underway to achieve a system of reliable, valid, and nationally available performance assessments—from a teacher’s point of entry through the development of accomplished teaching. Such a system would create a more useful and more common standard for the profession, just as national assessments do in fields such as nursing, engineering, accounting, medicine, and other skilled professions. A system of performance assessments could also leverage improvements in practice and professional learning opportunities.

As this paper details, some states have already begun to develop and implement standardized assessments of teacher performance that more accurately gauge the classroom effectiveness of beginning teachers, and a group of 20 states has joined together to build on these efforts to create a common tool for assessing novices. In addition, most states now recognize the National Board Certification program, which identifies veteran accomplished teachers who are more effective in developing student learning. The best practices from these initiatives can support a continuum across the teaching career for identifying and supporting stronger teaching and making more grounded personnel decisions based on a common, comprehensive set of standards that can be adopted nationwide to ensure that only the most well-prepared and effective teachers are instructing our public schools students.

In addition to raising the bar for teacher preparation and professional development, nationally available performance assessments at the points of the initial license, the professional license—usually about three years into the profession, just prior to tenure—and advanced certification could reflect the greater commonality in student expectations reflected in the so-called Common Core standards already adopted in more than 30 states. Such a system could also solve some of the problems created by the current Byzantine set of different licensing requirements across the 50 states and help create a national teacher labor market that supports mobility from states with surpluses to those with shortages while enhancing teacher quality.

A reliable and valid system of performance assessments based on common standards would provide consistency in gauging teacher effectiveness, help track educational progress, flag areas of need, and anchor a continuum of performance

throughout a teaching career. Such a system could also be used to establish standards for a National Teacher License that would allow mobility across states, ensure school districts that a new hire meets the requirements necessary to become an effective teacher who can advance student learning, and enable districts to identify and recruit the most able teachers to the most needy schools.

Why performance assessments are needed

In a recent report for the Center for American Progress, policy analyst Edward Crowe outlined a new accountability system for teacher education, one designed to redirect attention to the things that matter most: “Whether or not K-12 students are learning, how well teachers have developed the classroom teaching skills to be effective with their students, a graduate’s commitment to teaching as a professional career, feedback from graduates and employers, and high-quality tests of teacher knowledge and skills that are tied to classroom teaching performance and K-12 student learning.”³

Crowe notes that new assessments are needed to tell whether teacher education graduates have developed the classroom teaching skills to be effective with their students because current teacher tests don’t directly measure what teachers *do* in the classroom, and they don’t indicate how well teachers will do in the classroom. As a National Research Council report observed, most teacher licensure tests “are not constructed to predict the degree of teaching success a beginning teacher will demonstrate,” and studies suggest that they indeed do not.⁴

In nearly all states, teachers have to pass at least three tests—generally multiple-choice tests of basic skills, subject matter, and teaching knowledge—in order to become licensed, even though these are not strongly related to their ultimate success in the classroom. With individual states creating their own requirements, Crowe notes, “states have created a crazy quilt of... assessments that add up to 1,100 different tests” across the country.⁵ A great deal of money and energy is spent on developing tests that have little value in separating out teachers who are effective from those who are not. Furthermore, in many cases these tests evaluate teacher knowledge *before* they enter or complete teacher education, and hence are an inadequate tool for teacher education accountability.

Performance assessments that measure what teachers actually do in the classroom, and which have been found to be related to later teacher effectiveness, are a much more potent tool for evaluating teachers’ competence and readiness, as well as for supporting needed changes in teacher education.

Aside from teacher tests, there is increasing interest in measuring teachers' contributions by directly examining student achievement gains, and there are a number of efforts underway to create systems that incorporate value-added methods for examining student learning gains into teacher evaluation. As I describe in more detail later in this report, value-added methods have proved valuable for examining the potential influences on teacher effectiveness of teacher preparation programs, professional development programs, and various kinds of evaluation systems.

Yet, as Harvard University economics professor Thomas Kane pointed out in recent Senate testimony, these measures have been subject to concerns about their volatility at the individual teacher level and the possibility they could foster teaching toward narrow tests, as well as the fact that they are not available for about three-fourths of all teachers.⁶ This volatility, which is greater than that associated with observational measures, is due to the fact that the score gains measure more than the influence of the teacher—even when statistical methods are used to control for other factors, such as student characteristics, home and school resources, and the influence of other teachers, tutors, and parents on learning.

Furthermore, since most experts agree that at least three years of data about a given teacher are necessary to achieve a modicum of stability, the direct use of student test score data to evaluate teachers does not help inform judgments about new entrants to the profession. Yet in order to protect students, governments must make judgments about whether professionals are well enough prepared to practice safely and competently as soon as they enter the profession.

In his current work with the Gates Foundation, Kane and his colleagues are evaluating how certain kinds of classroom observations and videotapes of teaching, teacher reflections, content pedagogical assessments, and student and teacher feedback are related to measures of teacher effectiveness based on student achievement gains on both traditional tests and more intellectually challenging open-ended measures. This strategy aims to validate a range of teacher evaluation tools against value-added student learning gains using more than one way to look at student learning.

Economists Jonah Rockoff and Cecilia Sperroni have similarly noted that “value-added measures of effectiveness are noisy and can be biased if some teachers are persistently given students that are harder to teach in ways that administrative data do not measure. Thus, using other information may achieve more stability and accuracy in teacher evaluations.”⁷ By “noisy,” they mean unstable from year to year

and prone to error. They note that observation-based teaching evaluations, especially standards-based evaluations that carefully measure specific dimensions of teaching, have been found to be significantly related to student achievement gains. What's more, when these observations are used for feedback and coaching, they can help teachers develop greater effectiveness.⁸

In their work, Rockoff and Speroni confirmed that mentor teachers' evaluations were significant predictors of beginning teachers' current and subsequent value-added effectiveness. But they also found that the application of standards could vary significantly across evaluators. Thus, efforts to create more consistency in evaluating teacher performance are critical if performance is to be a central measure of teacher effectiveness.

As described below, structured teacher performance assessments address these needs. They evaluate directly what teachers do in the classroom, and they often incorporate contextualized evidence of student learning that is linked to evidence of the associated teaching efforts. Such assessments have been found to be stronger predictors of teachers' contributions to student learning gains than traditional teacher tests. They are also more reliably scored than most on-the-job classroom observations, and can be used for new teachers about to enter the profession, as well as long-time veterans. With recent advances, a continuum of such assessments can be constructed from initial entry to recognition of advanced expertise.

The last two decades have marked the emergence of professional standards for teaching stimulated in large part by the view that heightened expectations for student learning can be accomplished only by greater expectations for teaching quality. Those standards include those for National Board Certification and, more recently, for assessment of beginning teachers. Below we address each in turn.

National Board Certification

A standards-based approach to assessing teachers was initially developed and made systematic through the work of the National Board for Professional Teaching Standards. The board, founded in 1987, was the first body to involve expert teachers and researchers in developing standards for accomplished teaching. The standards outline what accomplished teaching looks like in more than 30 teaching areas defined by subject area and developmental level of students.

The National Board then developed an assessment that assembles evidence of teachers' practice and performance in a portfolio that includes videotapes of teaching accompanied by commentary, lesson plans, and evidence of student learning over time linked to evidence about the teachers' work with individual students. These pieces of evidence are scored reliably by trained raters who are knowledgeable in the same teaching field, using rubrics that define critical dimensions of teaching as the basis of the evaluation.

Designed to identify experienced and accomplished teachers, the National Board Certification is used in at least 32 states and more than 500 districts as the basis for teacher evaluation, salary increases, and other forms of teacher recognition, such as the selection of mentor or lead teachers. A number of states also provide license reciprocity on the basis of National Board status, and 28 use certification status as a proxy for license renewal. California offers a \$20,000 bonus, paid over four years, to board-certified teachers who teach in high-need schools, which has helped to distribute these accomplished teachers to students who need them.⁹

A number of recent studies have found that the National Board Certification assessment process distinguishes among teachers who are more and less effective in raising student achievement. Large-scale studies in Florida and North Carolina found that, controlling for a host of other student and teacher characteristics, students made significantly greater gains if their teachers were National Board Certified.¹⁰ Similarly, researchers in the Los Angeles Unified School District found that the positive effects of board-certified teachers grew even larger when examined using the stronger methodology of randomized assignment of classrooms to teachers.¹¹

In addition, smaller studies documenting positive influences of NBC teachers on their students' achievement have delved deeper to show how the practices of board-certified teachers differ from those of teachers who attempted but failed the assessment.¹² In particular, these studies show how board-certified teachers foster deeper understanding in their instructional design and classroom assignments.

One study with more mixed results found that National Board-certified teachers appeared more effective than others in some grade levels and subject areas, and on some tests but not in others.¹³ Another small study found positive but modest results that did not reach the level of statistical significance.¹⁴ To strengthen its ability to evaluate teachers' effectiveness, the National Board has just com-

pleted a study that will guide the incorporation of additional evidence of student learning into its assessments.¹⁵

Assessment of beginning teachers

National Board Certification, however, is reserved for experienced teachers. Until recently, there was no comparable assessment for beginning teachers that could evaluate who is ready to teach and likely to be effective.

Following on the work of the National Board, a consortium of states, working under the auspices of the Council of Chief State School Officers, created related standards for beginning teacher licensing that reflect professional teaching standards and incorporate student learning standards. More than 40 states have now adopted these standards into their licensing systems. In some states, teacher performance assessments for new teachers, modeled after the National Board assessments, have been developed for use either in teacher education, as a basis for the initial licensing recommendation (California, Colorado, Kentucky, Oregon), or in the teacher induction period, as a basis for moving from a probationary to a professional license (Connecticut).

These assessments require teachers to document their plans and teaching for a unit of instruction, videotape and analyze their teaching, and collect and evaluate evidence of student learning. All of these pieces of evidence are then assembled and evaluated by highly trained raters who score them in a consistent manner against specific criteria that reflect standards of best practice. As detailed in this paper, these assessments have been found to measure teacher effectiveness and can be used to help teachers develop greater effectiveness. Participation in these assessments has been found to support learning both for teachers who are being evaluated, and for educators who are trained to serve as evaluators.

For many years, Connecticut required beginning teachers to pass a Beginning Educator Support and Training, or BEST, performance assessment in their second or third year of teaching before they could be granted a professional license. The BEST assessment requirement was recently suspended because the state could not adequately fund the mentoring component, but while in place, beginning teachers' ratings on the Connecticut BEST assessment were found to significantly predict their students' value-added achievement on state reading tests.¹⁶ Using a sophisticated statistical technique called hierarchical linear modeling, a recent

study found that a one-unit increase in the BEST portfolio score was associated with a student score increase on the Degrees of Reading Power test equivalent to nearly half of students' average gain within a school year. In other words, students taught by a teacher who scored a "3" on the assessment's four-point scale gained about 40 percent more in reading during the course of the school year than students of a teacher who scored a "2" on the assessment.

Furthermore, the study found that the BEST portfolio scores were the only teacher characteristic that predicted student gains. Student gains were not significantly related to teachers' Praxis scores (a more traditional, primarily multiple-choice teacher test), the prestige of their pre-service institution, their race, ethnicity, gender, type of district, or participation in a mentoring program.

Meanwhile, in California, all teacher education programs—both traditional and alternative—must evaluate their candidates' teaching competence with a standardized performance assessment approved by the California Commission on Teacher Credentialing. The assessment is used to determine the initial teaching license, and the data are aggregated for use in program improvement and to inform state and national accreditation. The legislature's intent in enacting this requirement is to improve the quality of California teachers by holding programs accountable for the competence of their graduates.¹⁷

Built on the advances made by the National Board and Connecticut assessments, the Performance Assessment for California Teachers has emerged from this reform. Launched in 2002 and now used by 33 teacher education programs in the state—including district and university-based internship models, and a credentialing program run by a successful charter school, High Tech High of San Diego—PACT has been shown to be reliable, valid, and a strong lever for improving both teacher competence and program quality. Like the Connecticut BEST assessment, a preliminary validity study of PACT also found that teachers' scores on the assessment are positively associated with their value-added effectiveness when they later become full-time teachers.¹⁸

Significantly, PACT, which is taken after California's basic skills and subject matter tests, does not seem to pose additional barriers for aspiring teachers of color. Early validation studies of PACT have found no disparities in the outcomes of the assessment by candidate race or ethnicity, in contrast to many other teacher tests that have greatly disparate pass rates that have often reduced the diversity of the teaching force.¹⁹

Prospects for a national system of Teacher Performance Assessments

Based on this work, 20 states have recently joined together as a Teacher Performance Assessment Consortium under the auspices of the American Association of Colleges of Teacher Education, or AACTE, and the Council of Chief State School Officers, or CCSSO, to create a common initial licensing assessment that can be used nationwide to make preparation and licensing performance based, as well as predictive of teacher effectiveness. This assessment is built upon the model designed by the California PACT consortium.

A more advanced version of the assessment will also be developed for use at the point of the professional license—typically after the three-year probationary period—and to guide the mentoring process during the induction period. Success at this juncture could be linked with additional compensation in a state or district with a career ladder program.

The aforementioned assessments, which are subject specific, are grounded in model-teaching standards in the disciplines, which are themselves linked to common standards for student learning developed by CCSSO's Interstate New Teacher Support and Assessment Consortium, or INTASC.

Supported by foundation and state funds, and integrated into many states' proposals for the federal Race to the Top education grant program, the Teacher Performance Assessment Consortium has already completed the design of the assessment in the initial licensing areas and will pilot test these assessments in 2010-11 while completing the design of the remaining licensing areas. A larger pilot program is scheduled for the 2011-12 school year, and full scale-up program is slated for 2012-13 in "fast-track" states, including Massachusetts, Minnesota, New York, Tennessee, and Washington, all of which have already adopted the Teacher Performance Assessment as a requirement for licensing. Other states will bring the Teacher Performance Assessment on-line in the subsequent two years.

During this period of time, a subset of states will also develop the “Tier 2” assessment to be used at the point of transition from a probationary license to a professional license, as in Connecticut. (Tier 2 is the term given the second level of licensing in states that require the completion of induction requirements before initially licensed teachers are granted a longer-term professional license.) This assessment will be similar in design, but will focus less on teacher planning and more on evidence of student learning, and the capacity to use student learning results in designing instruction.

In a number of states, a continuum of teacher performance assessments is envisioned, as veteran teachers who are being evaluated for mentor or lead teacher status or for the receipt of higher levels of compensation are assessed through National Board Certification. By 2015, a national system of teacher performance assessments will be available for use in policy decisions, ranging from initial licensing to professional licensure and advanced certification.

This set of assessments can be used not only for personnel decision making over the course of the teaching career, but also for guiding teacher development and for evaluating and improving teacher education, mentoring, and professional development programs, as described in figure 1.

How performance assessments can help teachers improve their practice

Developing teacher effectiveness is as important as measuring it. Many studies have concluded that teachers’ participation in standards-based performance assessments can help teachers improve their practice. Teachers who have gone through National Board Certification, for example, note that the process of analyzing their own and their students’ work in light of professional standards helps them better assess student learning and evaluate the effects of their own actions. They also have to adopt new practices that are called for in the standards and assessments, such as engaging students in writing multiple drafts of papers or conducting science inquiries.²⁰

In addition, teachers reported that going through the board certification process caused them to improve their subject matter knowledge, design and delivery of instruction, classroom management, and evaluation of and support for student learning. A number of studies have documented that these changes do indeed

occur.²¹ In particular, teachers use new teaching strategies, pay more attention to student learning, and use assessments to change their practice to a much greater degree after they have gone through the assessment process.

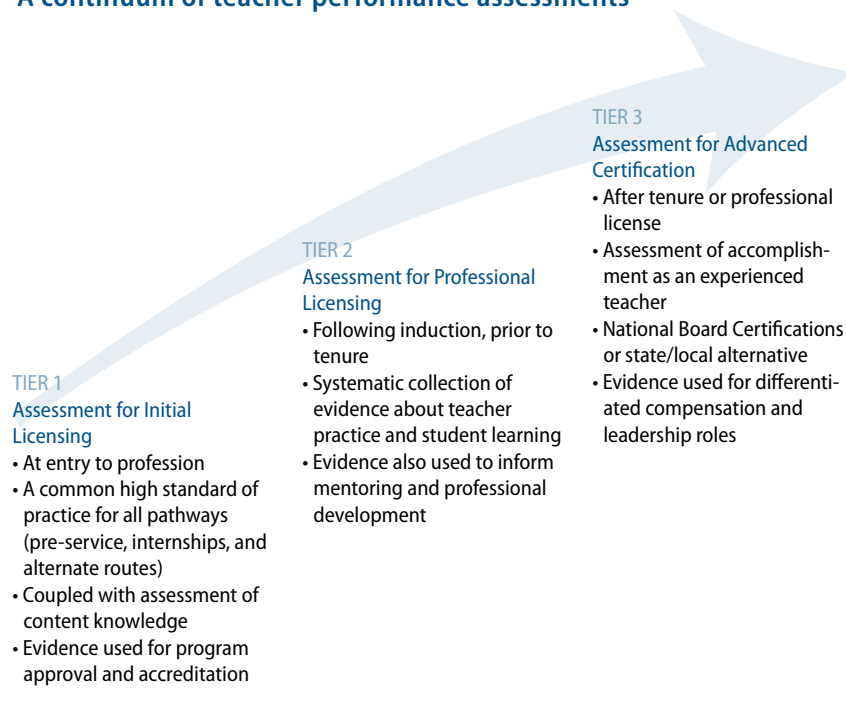
It is not unusual for National Board participants to say that they have learned more about teaching from their participation in the assessments than they have learned from any other professional development experience.²² Board-certified teacher David Haynes' statement is typical of many:

*Completing the portfolio for the Early Adolescence/Generalist Certification was, quite simply, the single most powerful professional development experience of my career. Never before have I thought so deeply about what I do with children, and why I do it. I looked critically at my practice, judging it against a set of high and rigorous standards. Often in daily work, I found myself rethinking my goals, correcting my course, moving in new directions. I am not the same teacher as I was before the assessment, and my experience seems to be typical.*²³

Performance assessments for beginning teachers also help novices improve their practice. A beginning teacher in Connecticut who participated in the BEST assessment described the power of the process, which required him to plan and teach a unit and to reflect daily on the day's lesson, considering how the lesson met the needs of each student and what should be changed in the next day's plans. He noted:

Although I was the reflective type anyway, it made me go a step further. I would have to say, okay, this is how I'm going to do it differently. It made more of an impact on my teaching and was more beneficial to me than just one lesson in

FIGURE 1
A continuum of teacher performance assessments



which you state what you're going to do...The process makes you think about your teaching and reflect on your teaching. And I think that's necessary to become an effective teacher.

Research on the PACT assessment used in California teacher education programs has found similar effects on candidates' learning. PACT assessments require student teachers or interns to:

- Plan and teach a week-long unit of instruction mapped to the state standards
- Reflect daily on the lesson they've just taught and revise plans for the next day
- Analyze and provide commentaries of videotapes of themselves teaching
- Collect and analyze evidence of student learning
- Reflect on what worked, what didn't, and why
- Project what they would do differently in a future set of lessons

Teaching candidates must show how they take into account students' prior knowledge and experiences in their planning. Adaptations for English language learners and for students with special needs must be incorporated into plans and instruction. Analyses of student outcomes are part of the evaluation of teaching. (See table 1.)

Some of these are things that many experienced teachers have never learned to do, but that have been found to be critical for student learning. Many teachers, for example, have learned to run through the curriculum, or get through the chapters of the textbook, without ever taking stock of what students understand in order to change their approach or re-teach concepts that weren't fully learned. Relatively few teachers have learned to analyze the learning outcomes of their students in a nuanced way that would guide their work with individual students and their broader curriculum planning.

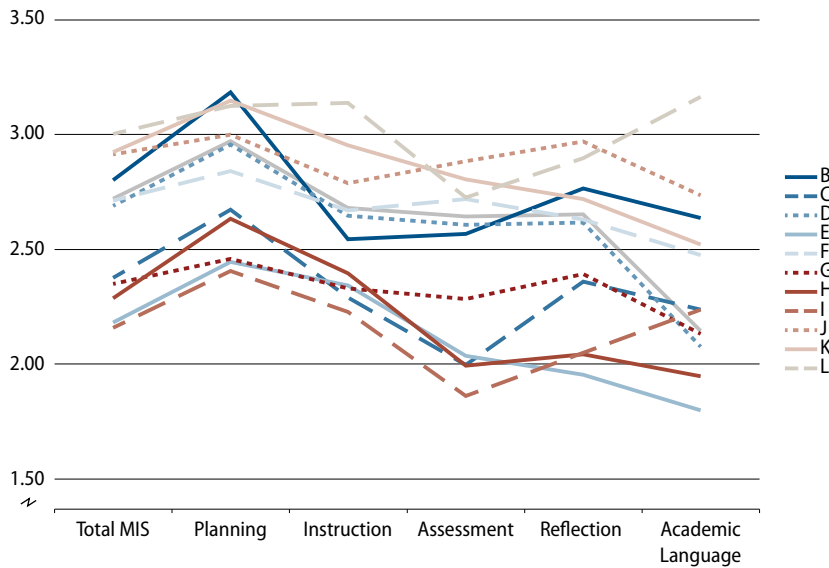
The requirement that beginning teachers evaluate student learning daily to adjust their plans and to evaluate student learning growth changes their understanding of teaching and their practice. For example, prospective teachers have noted after completing the PACT:

For me, the most valuable thing was the sequencing of the lessons, teaching the lesson and evaluating what the kids were getting, what they weren't getting, and having that be reflected in my next lesson...the "teach-assess-teach-assess-teach-assess" process. And so you're constantly changing—you may have a plan or a framework, but you know that that has to be flexible, based on what the children learn that day.²⁴

Table 1: Overview of mathematics teaching event

Teaching event task	What to do	What to submit
1. Context for learning	<ul style="list-style-type: none"> • Provide relevant information about your instructional context and your students as learners of mathematics. 	<ul style="list-style-type: none"> • Context form • Context commentary
2. Planning instruction and assessment	<ul style="list-style-type: none"> • Select a learning segment of three to five hours of instruction that develops students’ mathematical knowledge by developing a balance of procedural fluency, conceptual understanding, and mathematical reasoning. It should also foster positive student dispositions toward mathematics. • Create an instruction and assessment plan for the learning segment and write lesson plans. • Provide accommodations for English learners, students with disabilities, and any other students with specific needs. • Write a commentary that explains your thinking in writing the plans. • Record daily reflections about what happened as you taught, and adapt your plans accordingly. 	<ul style="list-style-type: none"> • Lesson plans for learning segment • Instructional materials • Planning commentary
3. Instructing students and supporting learning	<ul style="list-style-type: none"> • Review your plans and prepare to videotape your class. Identify opportunities for students to understand mathematical concepts, procedures, and reasoning. • Videotape the lesson(s) you have identified. • Review the videotape to identify one or two video clips portraying the required features of your teaching. The total running time should not exceed 20 minutes. • Write a commentary that analyzes your teaching and your students’ learning in the video clip(s). 	<ul style="list-style-type: none"> • Video clip(s) • Video label form • Instruction commentary
4. Assessing student learning	<ul style="list-style-type: none"> • Select one student assessment from the learning segment and analyze student work. • Identify three student work samples that illustrate class trends in what students did and did not understand. • Write a commentary that analyzes the extent to which the class met the standards/objectives, analyzes the individual learning of two students represented in the work samples, describes feedback to students, and identifies next steps in instruction. 	<ul style="list-style-type: none"> • Student work samples • Evaluative criteria or rubric • Assessment commentary
5. Reflecting on teaching	<ul style="list-style-type: none"> • Provide your daily reflections. • Write a commentary about what you learned from teaching this learning segment. 	<ul style="list-style-type: none"> • Daily reflections • Reflective commentary

FIGURE 2
PACT scores by institution



The assessment piece [of the PACT] was good. Like for the math... I really got into that—tallied [the data] all up in Excel, and made a graph. And that was kind of fun. It was like, “Oh, I could probably do this more often”... you know, really digging into their work and looking for what was going on. I should make that more of a habit next year than I have this year, now that I know.²⁵

Studies show that beginning teachers are capable of

assessment in their actual classroom practice. Research on student teachers who had completed the Performance Assessment for California Teachers found that pre-service teachers did change their teaching practices as a consequence of their experiences with the performance assessment,²⁶ and first-year teachers reported continued influences of the assessment on their teaching.²⁷

One beginning teacher followed through the process, we will call her Joy, observed that there were many things about the assessment that strengthened her practice. Analyzing the teaching videotape forced her to reflect on her teaching “in a different, much deeper way,” while the need to assess student progress “helped me to get focused and ... to see that there’s ... a need for continuity in the lesson, [and] also to look at where [the students] are ... at the end of the day and maybe change things a little bit to find out where they need to go the next day.”

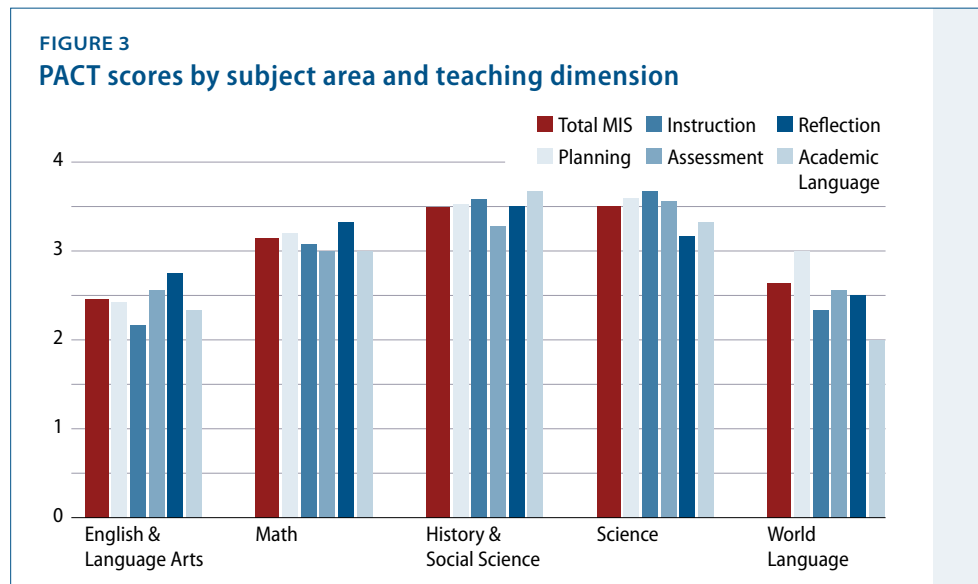
When asked about her students’ backgrounds and skills levels in an initial interview, Joy was at a loss and had no specific information. Later, after having completed the task requiring that specific information for the PACT, Joy underscored the value of learning about her students in this way:

I know that ... this coming school year, I'm going to get out the sheets the parents fill out. You know, "How old is the child? And what is their nationality? When did they come to the United States? And did they have other brothers and sisters? What is their background?" You know, it really helps you to understand your class and each child much better. I'm going to make that a real priority, where I really wouldn't have thought about doing that... I learned a lot from that.²⁸

Of the nine skills Joy reported learning from the PACT experience, only three were skills she had been taught in her teacher education program. Planning and teaching full curriculum units in both literacy and mathematics and evaluating her own practice and her students' learning would not normally have happened in her programs' design for a student teaching experience were it not for the PACT requirements. This suggests not only how candidates can learn from an assessment, but also how a well-designed performance assessment for licensing can leverage major changes in teacher education.

Improving preparation through the use of performance assessments

There are a number of ways that performance assessments for licensing influence teacher preparation programs beyond the effects on individual candidates. PACT assessments are scored by trained raters—faculty members and supervisors as well as cooperating teachers and principals in schools that help train teachers—whose ratings are further moderated and audited to produce highly reliable and valid teacher performance evaluations. Programs receive detailed, aggregated data on all of their candidates by program area and dimensions of teaching (see figures 2 and 3), and use the data to improve curriculum and program designs. The aggregated data will ultimately be used for program accredi-



tation to provide a basis for deciding which programs should be encouraged, improved, or closed if they cannot improve enough to enable most of their candidates to demonstrate that they can teach.

The aggregated scores are helpful for faculty looking for patterns of performance, but the act of scoring is itself also educative. As is true for teachers of elementary and secondary school students, there is great power in looking closely at student work, evaluating whether and how it meets challenging standards, and deliberating collectively about how to improve curriculum and teaching to ensure greater success.²⁹ Professional development for personnel in schools hosting aspiring teachers, as in colleges of education, can benefit greatly from this kind of organizing focus and energy.

Faculty, supervisors, cooperating teachers, and other educators who score these portfolios use standardized rubrics in moderated sessions following training with an audit procedure to calibrate standards. These participants describe a range of benefits for teacher preparation from the scoring processes. Here are some examples:

This [scoring] experience... has forced me to revisit the question of what really matters in the assessment of teachers, which—in turn—means revisiting the question of what really matters in the preparation of teachers.

— A teacher education faculty member

[The scoring process] forces you to be clear about “good teaching;” what it looks like, sounds like. It enables you to look at your own practice critically, with new eyes.

— A cooperating teacher

As an induction program coordinator, I have a much clearer picture of what credential holders will bring to us and of what they’ll be required to do. We can build on this.

— An induction program coordinator

One of the more powerful things about a performance assessment of this kind is that it requires teacher candidates to pull together all the many things they are supposed to be learning in courses and clinical experiences—how to diagnose student learning, plan in response to standards, manage and revise instruction, and evaluate the outcomes for student understanding—into a single coherent teaching event. It may be the first and only time in a program that candidates and their instructors can see

whether they indeed understand and can apply what they are supposed to be learning. Many programs have to make major changes to accomplish this, by integrating areas of knowledge, reducing fragmentation among courses and clinical experiences, increasing applications to practice, and paying more attention to areas that have traditionally been underdeveloped in teachers' repertoires.

In California, for example, most programs found that their candidates scored least well on the assessment dimension and on the academic language dimension, which is a set of criteria evaluating how well teachers both develop English language proficiency for their students and help them acquire and use academic language in the discipline. To evaluate their candidates' abilities, some large programs draw a random sample of candidates for a joint mock scoring session. This allows faculty to look at common evidence of teaching together and discuss it. As one program leader noted, seeing the candidates' plans, videotapes of teaching, and commentary is often an eye-opening experience for faculty, often leading to serious curriculum changes:

The persuasive piece was once they saw the student work. I mean, where a few people kind of went, 'Whoa.' I teach this in my class and I'm not seeing it... looking at the student work from the mock scoring there was that 'ah hah' moment where [it was clear that] our candidates didn't know much about (academic language)... [One professor] changed her entire series of assignments... to better reflect what the holes in the data [showed]—also to incorporate more clearly the notion of academic language and mathematics. She literally rewrote everything related to that assignment because it was so compelling to her, the data... and seeing the student work.³⁰

In this very large program, analysis of the PACT data led to more comprehensive and collaborative approaches across the program. Another faculty member noted:

Frankly, that course [on language acquisition] has been marginalized for a long time. It's sort of a stand-alone course. Other faculty members don't really pay much attention to language acquisition in their methods courses. ... So, I think everybody recognized we have to do more across courses.³¹

This problem was subsequently addressed through new assessments in the program, as well as a professional development series in each of the department meetings and at the faculty retreats. Analyses of PACT scores and actual portfolios

were used as a prelude to curriculum conversations. Faculty attributed subsequent improvements in PACT scores to these changes.

Another program described a similar process of involving school- and university-based faculty in examining the PACT portfolios as a way to evaluate what aspects of the program were producing the desired outcomes. A faculty leader observed:

I think there are two domains that are problematic for everybody, and that's assessment and academic language. So we have tackled both of them throughout this process, with great attention. I think we've been improving a lot in both... In the area of assessment it seems like [candidates] are better at kind of looking at assessment as PACT measured it in the planning, which is basically kind of a diagnostic assessment. They're less effective at doing this kind of loop, where you think "what progress is made in student learning over time, and what can you get from it?" So we've been working on that and, in fact, using our inquiry projects as the instructional intervention for us to enhance that [skill].³²

As the program director at yet another program noted, the PACT data stimulated much more change than other accountability policies. According to the program director:

... (W)hen you have these 'standards', these documents—it's very easy to say, 'We do that; we do that. Here, here's my syllabus. Here are the activities I do. We address this. We prepare them for that.' And it's very easy to do that... to write to those kinds of documents... It's a very different thing when you have your students engaging in an assessment, and then you're seeing the results of that assessment. It starts to challenge in particular ways what you do. So it wasn't until we were really working with the (PACT) data that we felt more threatened... The value comes in when they're sitting down during the scoring process. If you're sitting down together in a room with folks looking at evidence of teaching, it's a very powerful way of understanding teaching practice and helping you look at your own practice."³³

The process in this program resulted in more shared knowledge and competence among teacher educators, since “in preparing (candidates to do PACT) and in scoring their work, every scorer and every person preparing them has to have some understanding of how you support students with special needs or how you ensure that academic language is addressed and that children for whom English is a second language are supported throughout their teaching.” The integration of

knowledge that teachers have to do in practice was now required of teacher educators as well. As the program director observed:

“[Previously], most people knew their own practice. They knew what they would do when they were supervising their students. They would know what the candidates were doing within their own courses. And that was the extent of it... [But now] in part because people have learned more about teaching practice on a number of different levels, outside their area of specialty, because they have learned more about the program as a whole, we have far more people who are able to participate and think about the ways in which we improve our program.”³⁴

The upshot: When assessments both predict teacher effectiveness and support individual and institutional learning, they can help to create an engine for stimulating greater teacher effectiveness in the system as a whole.

To be sure, though, the adoption of this kind of performance assessment system is complex, and it will likely require a greater allocation of resources to teacher preparation from both institutions and government funders than is currently allocated in many states’ systems. Many California colleges, universities, and alternative programs have been making greater investments in their teacher education efforts to meet the higher standards posted by the performance assessments for well-guided clinical training and coursework in previously neglected areas, such as assessment and the teaching of English learners.

Much work needs to be done to build opportunities to learn into the curriculum, to prepare school- and university-based teacher educators and teacher candidates to understand and implement the assessments, and to motivate teacher educators to spend the time needed to examine and respond to the assessment data on their students.³⁵

California institutions have estimated that the cost of implementing and scoring the performance assessment and maintaining procedures for reliability and validity checks could cost on average about \$400 per candidate. Ideally, states would organize support for this implementation and scoring process, and would seek to assume some of the costs that might otherwise be charged to candidates in fees. Yet many states are cutting support for higher education, including teacher education programs, during the current economic recession. Ultimately, securing funding and programmatic supports for this work will require both state and federal initiatives—for example, through Title II of the Higher Education Act, which cov-

ers teacher education accountability and program initiatives—along with coordinated efforts on the part of program providers to support an implementation process that allows programs to learn from one another rather than reinventing the wheel in each locale.

Policy recommendations and conclusion

Teacher Performance Assessments for initial and professional licensing can support more rigorous evaluation and more purposeful development across a variety of routes into teaching, and can increase the consistency with which teacher licensure decisions are made across states. The assessments will provide information that states can use, not only to issue more meaningful teacher licenses, but also to inform teacher quality initiatives, make accreditation decisions, and plan teacher induction and in-service development. Used in conjunction with other measures, such as tests of teachers' subject matter knowledge, the assessments can support teacher quality improvements in several ways:

States can use Teacher Performance Assessments to improve the consistency and quality of data on beginning teacher effectiveness and anchor a continuum of performance assessments throughout the teaching career.

- Teacher education programs can use TPA data to flag program needs, guide improvements, and track progress.
- States, school districts, schools, and teacher development programs can use the assessments to provide an evidence-based methodology for making systematic decisions about recruitment, employment, professional development, and career development, as well as an outcome database that can be used by school districts to manage, analyze, and report data about teacher outcomes, and to track performance across the continuum of teachers' careers.
- States and accreditors can use TPA outcome data as information for the accreditation process to leverage significant improvements in preparation programs, especially if accreditors adopt an expectation that programs must show a specific level of performance, for example, 70 percent of candidates passing the performance assessment, in order to maintain approval.

- Induction programs can use the assessments to guide more effective mentoring for beginning teachers, providing information to mentors as candidates enter teaching, and guiding the mentoring process toward the assessment used at the end of the probationary period as well.
- States and the federal government can use nationally available teacher performance assessments, along with National Board Certification, to create a portable license that will facilitate teacher mobility across states. High scorers on these performance assessments could be granted a National Teacher License that would allow them mobility across states, and might make them eligible for incentives to attract effective teachers to high-need schools.
- A system of nationally available teacher performance assessments would allow states, school districts, and preparation programs to share a common framework for defining and measuring a set of core teaching skills that form a valid and robust vision of teacher effectiveness, reflecting both teacher practices and student learning. The current state consortium initiative that is creating such a system, supported by the Council of Chief State School Officers and the American Association of Colleges for Teacher Education, will team with one or more testing companies to support common systems of administering the assessments across the states.

As states use such assessments to inform teacher licensure, recruitment, induction, retention, and recognition, they will move toward a national standard of practice that advances student learning. The assessments can support efforts to evaluate and strengthen the connection between teacher performance and student outcomes with valid and reliable data that can also be used to guide pre-service and in-service training.

The hope is that the kind of assessments described in this report can contribute to the development of a more coherent and comprehensive national policy environment for teacher licensure, recruitment, and in-service evaluation, and ultimately to a more effective national agenda for improvement of teacher quality.

Endnotes

- 1 Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, "Teacher credentials and student achievement in high school: A Cross subject analysis with student fixed effects," Working paper series, Number 13617, (Stanford: National Bureau of Economic Research, 2007)
- 2 For a description of teacher policies in high-achieving nations, see Linda Darling-Hammond, *The Flat World and Education*. (New York: Teachers College Press, 2010)
- 3 Edward Crowe, "Measuring What Matters: A Stronger Accountability Model for Teacher Education" (Washington: Center for American Progress, 2010).
- 4 Karen J. Mitchell, David Z. Robinson, Barbara S. Plake, and Kaeli T. Knowles, "Testing Teacher Candidates: The Role of Licensure Tests in Improving Teacher Quality" (Washington, DC: National Academy Press, 2001).
- 5 Crowe, "Measuring What Matters," p. 7.
- 6 Thomas Kane, Testimony before the U.S. Senate Health, Education, Labor and Pensions Committee, April 15, 2010.
- 7 Jonah Rockoff and Cecilia Speroni (2010). *Subjective and Objective Evaluations of Teacher Effectiveness* (New York: Columbia University, 2010) p. 1.
- 8 Anthony Milanowski, "The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence From Cincinnati," *Peabody Journal of Education* 79 (4) (2004): 33-53.
- 9 Daniel Humphrey, Julie Koppich, and Heather Hough, "Sharing the wealth: National Board Certified Teachers and the students who need them most," *Education Policy Analysis Archives* 13 (18) (2005). Retrieved June 10, 2008, from <http://epaa.asu.edu/epaa/v13n18/v13n18.pdf>.
- 10 Clotfelter, Ladd, and Vigdor, "Teacher Credentials and Student Achievement;" Charles Clotfelter, Helen Ladd, and Jacob Vigdor, "Teacher-Student Matching and the Assessment of Teacher Effectiveness," NBER Working Paper 11936 (National Bureau of Economic Research, 2006); L. Cavaluzzo, "Is National Board Certification an effective signal of teacher quality?" (Alexandria, VA: The CNA Corporation, 2004); D. Goldhaber & E. Anthony, "Can teacher quality be effectively assessed?" (Seattle, WA: University of Washington and the Urban Institute, 2005).
- 11 Cantrell, Fullerton, Kane, and Staiger (2007) found that students of NBC teachers outperformed those of teachers who had unsuccessfully attempted the certification process by 0.2 standard deviations, about twice the differential that they found between NBC teachers and unsuccessful applicants from a broader LAUSD sample not part of the randomized experiment, but analyzed with statistical controls. Their correlational findings – also finding positive effects – were comparable to those of the other studies noted above.
- 12 Lloyd Bond and others, "The Certification System of the National Board for Professional Teaching Standards: A Construct and Consequential Validity Study" (Greensboro, NC: Center for Educational Research and Evaluation, 2000); Tracy Smith and others, "An Examination of the Relationship of the Depth of Student Learning and National Board Certification Status" (Appalachian State University, 2005); Leslie Vandevoort, Audrey Amrein-Beardsley, and David Berliner, "National Board Certified teachers and their students' achievement," *Education Policy Analysis Archives* 12 (46) (2004).
- 13 Douglas Harris and Tim Sass, "The effects of NBPTS-certified teachers on student achievement" (2007), available at http://www.nbpts.org/userfiles/file/harris_sass_final_2007.pdf. This study found differential effects of National Board certified teachers on the two tests examined: the Florida Comprehensive Assessment Test and the SAT-9. It did not compare board certified teachers to those who had attempted certification unsuccessfully, which is the best comparison set for answering the question of whether the Board's process differentiates between more and less effective teachers.
- 14 William Sanders, James Ashton, and S. Paul Wright, "Comparison of the Effects of NBPTS Certified Teachers with Other Teachers on the Rate of Student Academic Progress" (2005), available at http://www.nbpts.org/UserFiles/File/SAS_final_report_Sanders.pdf. This study found effect sizes for National Board-certified teachers similar to those of other studies (about .05 to .07 in math), but most of the estimates were not statistically significant because of the small sample size. It did not compare board-certified teachers to those who had attempted certification unsuccessfully.
- 15 National Board for Professional Teaching Standards, "Teacher Evaluation in a New Era of School Improvement" (2010).
- 16 Mark Wilson and others, "Using Student Achievement Test Scores as Evidence of External Validity for Indicators of Teacher Quality: Connecticut's Beginning Educator Support and Training Program (Berkeley, CA: University of California at Berkeley, 2007), available at http://scale.stanford.edu/index.php?option=com_content&view=article&id=146:using-student-achievement-test-scores-as-evidence-of-external-validity-for-indicators-of-teacher-quality-connecticuts-beginning-educator-support-and-training-program&catid=59:latest-news&Itemid=107.
- 17 Charles Peck and Morva McDonald, "Creating 'Cultures of Evidence' in Teacher Education: "Context, Policy and Practice in Three High Data Use Programs" (Washington: Agency for Education Development, 2010).
- 18 Stephen P. Newton, "Predictive Validity of the Performance Assessment for California Teachers" (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010), available at <http://scale.stanford.edu/>.
- 19 Raymond Pecheone and Ruth Chung, "Evidence in Teacher Education: The Performance Assessment for California Teachers," *Journal of Teacher Education* 57 (1) (2006): 22-36.
- 20 Steven Athanases, "Teachers' Reports of the Effects of Preparing Portfolios of Literacy Instruction," *Elementary School Journal* 94 (4) (1994): 421-439.
- 21 Edward Chittenden and Jacqueline Jones, "An Observational Study of National Board Candidates as they Progress Through the Certification Process," Paper presented at the annual meeting of the American Educational Research Association, April 1997; Mistilina Sato, Ruth Chung Wei, and Linda Darling-Hammond, "Improving Teachers' Assessment Practices Through Professional Development: The Case of National Board Certification," *American Educational*

- Research Journal 45 (3) (2008): 669-700. S. Tracz, S/ Sienty, and S. Mata, "The Self-Reflection of Teachers Compiling Portfolios for National Certification: Work in Progress," Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, February 1994; S. Tracz and others, "Improvement in teaching skills: Perspectives from National Board for Professional Teaching Standards field test network candidates," Paper presented at the annual meeting of the American Educational Research Association, April 1995.
- 22 David Haynes, "One Teacher's Experience with National Board assessment," *Educational Leadership*, 52 (8) (1995): 58-60; Ann Bradley, "Pioneers in Professionalism," *Education Week* 13 (1994): 18-21; Areglado, N. "I Became Convinced: How a Certification Program Revitalized an Educator," *National Staff Development Council Winter* (1999): 35-37; M. Buday and J. Kelly, "National Board Certification and the Teaching Professions Commitment to Quality Assurance. *Phi Delta Kappan*, 78 (3) (1996): 215-219.
 - 23 Haynes, p. 60.
 - 24 Linda Darling-Hammond & Ruth Chung Wei, with Christy Marie Johnson, "Teacher Preparation and Teacher Learning: A Changing Policy Landscape," in Gary Sykes (ed.) *The Handbook of Education Policy Research* (Washington: American Education Research Association, 2009) p. 627.
 - 25 Ruth R. Chung, "Beyond Assessment: Performance Assessments in Teacher Education," *Teacher Education Quarterly* 35(1)(2008): 7-28, at p. 14.
 - 26 Chung, "Beyond Assessment." For data on candidate and program effects of teacher performance assessment, see also Ruth Chung Wei and Raymond Pecheone, "Teaching Performance Assessments as Summative Events and Educative Tools., in Mary Kennedy (ed.), *Teacher Assessment and Teacher Quality: A Handbook* (New York: Jossey-Bass, 2010).
 - 27 T. Sloan, L. Cavazos, and A. Lippincott, "A Holistic Approach to Assessing Teacher Competency: Can One Assessment Do it All? Paper presented at the Annual Meeting of the American Educational Research Association, April 2007.
 - 28 Chung, *Beyond Assessment*, p. 14.
 - 29 Linda Darling-Hammond, "Performance Counts: Assessment Systems that Support High-Quality Learning". (Washington: Council for Chief State School Officers, 2010).
 - 30 Peck and McDonald, "Creating 'Cultures of Evidence' in Teacher Education," p. 8, 9-10.
 - 31 Peck and McDonald, "Creating 'Cultures of Evidence' in Teacher Education," p. 10.
 - 32 Peck and McDonald, "Creating 'Cultures of Evidence' in Teacher Education," p. 16.
 - 33 Peck and McDonald, "Creating 'Cultures of Evidence' in Teacher Education," p. 22.
 - 34 Peck and McDonald, "Creating 'Cultures of Evidence' in Teacher Education," p. 23.
 - 35 For a thoughtful discussion of the issues that need to be addressed to implement these kinds of assessments in teacher education programs, see Charles Peck, Chrysan Gallucci, and Tine Sloan, "Negotiating implementation of high-stakes performance assessment policies in teacher education: From compliance to inquiry," *Journal of Teacher Education* 61 (3)(2010), available at <http://jte.sagepub.com/content/early/2010/03/12/0022487109354520>.

About the author

Linda Darling-Hammond is currently Charles E. Ducommun Professor of Education at Stanford University where she has founded and co-directs the Stanford Center for Opportunity Policy in Education. She is a former president of the American Educational Research Association and a member of the National Academy of Education.

Acknowledgements

The author wishes to express thanks to the Carnegie Corporation of New York, the Ford Foundation, and the Morgan Family Foundation for their support of the research underlying this report. In addition, a number of people offered very helpful ideas and assistance with this project: Robin Chait and Raegen Miller, Center for American Progress; Raymond Pecheone, Kendyll Stansbury, Ruth Chung Wei, and Nicole Merino, at Stanford University; and Cap Peck at University of Washington. Finally, for expert review and helpful suggestions, thanks to Bob Floden at Michigan State University and Kenneth Zeichner at the University of Washington.

The Center for American Progress thanks the Bill and Melinda Gates Foundation and the Joyce Foundation for generously providing support for this paper.

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.”

