The International Journal of Educational and Psychological Assessment
December 2009, Vol. 3

Recognizing and Enhancing Teacher Effectiveness

Linda Darling-Hammond
Stanford University, USA

Abstract

Efforts to improve the quality of teaching have begun to focus on ways to assess teacher effectiveness and to create systems of development and rewards that support more effective teaching. Policymakers and practitioners are seeking means to evaluate and recognize teacher effectiveness for the purposes of licensing, hiring, and granting tenure; for providing needed professional development; and for identifying expert teachers who can be recognized and rewarded. Some policy makers are also interested in tying compensation to judgments about teacher effectiveness, either by differentiating wages or by linking such judgments to additional responsibilities that carry additional stipends or salary. This paper draws upon research in outlining the issues associated with various approaches to ascertaining teacher effectiveness, and suggests a framework for policy systems that might prove productive in both identifying and developing more effective teachers and teaching.

Introduction

As nations pay increasing attention to educational outcomes, policymakers have undertaken a wide range of reforms to improve schools, ranging from new standards and tests to redesigned schools, new curricula, and new instructional strategies. One important lesson from these efforts has been the recurrent finding that teachers are the fulcrum that determines whether any school initiative tips toward success or failure. Every aspect of school reform -- the creation of more challenging curriculum, the use of ambitious assessments, the implementation of decentralized management, the invention of new model schools and programs -- depends on highly-skilled teachers.

Reformers have learned that successful programs or curricula cannot be transported from one school to another where teachers do not know how to use them well. Raising graduation requirements has proved to be of little use where there are not enough qualified teachers prepared to teach more advanced subjects well. Mandates for more math and science courses are badly implemented when there are chronic shortages of teachers prepared to teach these subjects. Course content is diluted and more students fail when teachers are not adequately prepared for the new courses and students they must teach. In the final analysis, there are no policies that can improve schools if the people in them are not armed with the knowledge and skills they need.

Furthermore, teachers need even more sophisticated abilities to teach 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 or difficulties. Clearly, 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.

An aspect of this transformation is developing means to evaluate and recognize teacher effectiveness throughout the career, for the purposes of licensing, hiring, and granting tenure; for providing needed professional development; and for recognizing expert teachers who can be recognized and rewarded. A goal of such recognition is to keep talented teachers in the profession and to identify those who can take on roles as mentors, coaches, and teacher leaders who develop curriculum and professional learning opportunities, who redesign schools, and who, in some cases, become principals. Some policymakers are also interested in tying compensation to judgments about teacher effectiveness, either by differentiating wages or by linking such judgments to additional responsibilities that carry additional stipends or salary. An integrated approach connects these goals with a professional development system into a career ladder.

In this paper, I draw on research in outlining the issues associated with various approaches to ascertaining teacher effectiveness, and I suggest a framework for policy systems that might prove productive in both identifying and *developing* more effective teachers and teaching. I draw a distinction between effective teachers and effective teaching that is important to consider if improvement in student learning is the ultimate goal.

Effective Teachers and Teaching

It is important to distinguish between the related but distinct ideas of teacher quality and teaching quality. *Teacher quality* might be thought of as the bundle of personal traits, skills, and understandings an individual brings to teaching, including dispositions to behave in certain ways. The traits desired of a teacher may vary depending on conceptions of and goals for education; thus, it might be more productive to think of teacher *qualities* that seem associated with what teachers are expected to be and do.

Research on teacher effectiveness, based on teacher ratings and student achievement gains, has found the following qualities important:

(1) strong general intelligence and verbal ability that help teachers organize and explain ideas, as well as to observe and think diagnostically;

(2) strong content knowledge – up to a threshold level that relates to what is to be taught;

(3) knowledge of how to teach others in that area (content pedagogy), in particular how to use hands-on learning techniques (e.g. lab work in science and manipulatives in mathematics) and how to develop higher-order thinking skills.

(4) an understanding of learners and their learning and development– including how to assess and scaffold learning, how to support students who have learning differences or difficulties, and how to support the learning of language and content for those who are not already proficient in the language of instruction.

(4) adaptive expertise that allow teachers to make judgments about what is likely to work in a given context in response to students' needs.[1]

Although less directly studied, most educators would include this list a set of dispositions to support learning for all students, to teach in a fair and unbiased manner, to be willing and able to adapt instruction to help students succeed, to strive to continue to learn and improve, and to be willing and able to collaborate

with other professionals and parents in the service of individual students and the school as a whole.

These qualities, supported by research on teaching, are embodied in the standards adopted by the National Board for Professional Teaching Standards and, at the beginning teacher level, by the states involved in the Interstate New Teacher Assessment and Support Consortium (INTASC), operating under the aegis of the Council of Chief State School Officers (CCSSO). This consortium of more than 30 states has taken a leading role in developing both new teacher standards and assessments and has led to the adoption of new licensing standards in most states. As these standards have been built into licensing and preparation requirements over the last decade, they have provided a means to develop a stronger foundation for effective teaching, making teacher qualifications a stronger predictor of teacher effectiveness.

Teaching quality has to do with strong instruction that enables a wide range of students to learn. Such instruction meets the demands of the discipline, the goals of instruction, and the needs of students in a particular context. Teaching quality is in part a function of teacher quality – teachers' knowledge, skills, and dispositions – but it is also strongly influenced by the context of instruction. Key to considerations of context are "fit" and teaching conditions. A "high-quality" teacher may not be able to offer high quality instruction in a context where there is a mismatch in terms of the demands of the situation and his or her knowledge and skills; for example, an able teacher asked to teach subject matter for which s/he is not prepared may teach poorly; a teacher who is prepared and effective at the high school level may be unable to teach small children; and a teacher who is able to teach high-ability students or affluent students well may be quite unable to teach students who struggle to learn or who do not have the resources at home that the teacher is accustomed to assuming are available. Thus, a high-quality teacher in one circumstance may not be a high-quality teacher for another.

A second major consideration in the quality of teaching has to do with the conditions for instruction. If high-quality teachers lack strong curriculum materials, necessary supplies and equipment, reasonable class sizes, and the opportunity to plan with other teachers to create both appropriate lessons and a coherent curriculum across grades and subject areas, the quality of teaching students experience may be suboptimal, even if the quality of teachers is high. Many conditions of teaching are out of the control of teachers and depend on the administrative and policy systems in which they work.

Strong teacher quality may heighten the probability of strong teaching quality, but does not guarantee it. Initiatives to develop teaching quality must consider not only how to identify, reward, and use teachers' skills and abilities but how to develop teaching contexts that enable good practice on the part of teachers. Hiring knowledgeable teachers but asking them to teach out of field, without high-quality curriculum or materials, and in isolation from their colleagues diminishes teaching quality and student learning. Thus, the policies that construct the teaching context must be addressed along with the qualities and roles of individual teachers.

Means for Identifying Effective Teaching for Policy Purposes

In recent years, there has been growing interest in moving beyond traditional measures of teacher qualifications – for example, a score on a paper-and-pencil test or completion of a preparation program before entry, or years of experience and degrees for in-service teachers – to evaluate teachers' actual performance and effectiveness as the basis for making decisions about hiring, tenure, licensing, compensation, and selection for leadership roles. The recent report of the No Child Left Behind (NCLB) Commission in the United States called for moving beyond the designation of teachers as "highly qualified" to an assessment of "highly effective" teachers based on their students' gains on state tests. Other recent U.S. proposals (for example, the TEACH 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.

Some state and local policymakers have sought to develop career ladders or other compensation plans that take into account various measures of teacher effectiveness for designating teachers for specific roles or rewards. These have included measures like National Board Certification and other performance-based evaluations, indicators like master's degrees and years of experience, and various measures of student learning. In addition, a few states have developed performance-based assessments for beginning teacher licensing as a means of determining effectiveness before teachers receive tenure or a professional license.

This paper reviews three categories of measures: 1) Evidence of teacher performance; 2) evidence of teacher knowledge, skills, and practices associated with student learning; and 3) evidence of student learning, including value-added student achievement test scores. Most career ladder or performance-based compensation plans that have survived to date use a combination of all of these measures, a point to which I return in the final section.

I discuss what is known in each category regarding both the validity of the measures and the influence of using certain measures or approaches on the improvement of teaching practice. The presumption underlying this discussion is that successful policies will seek to develop systems that both *assess* teacher effectiveness in valid ways and help to *develop* more effective teachers at both the individual and collective levels.

Evidence of Teacher Performance

There is growing evidence that some well-designed performance-based assessments of teaching detect aspects of teaching that are significantly related to teacher effectiveness, as measured by student achievement gains. These include standardized *teacher performance assessments* like those used for National Board Certification and for beginning teacher licensure in states like Connecticut and California, as well as *standards-based teacher evaluation systems* used in some local districts. The value of using such assessments is that they can both document broader aspects of teacher effectiveness and can be used to help teachers develop greater effectiveness, as participation in these assessments has been found to

support learning both for teachers who are being evaluated and educators who are trained to serve as evaluators.

Teacher Performance Assessments. A standards-based approach to assessing teachers was initially developed and made systematic through the work of the National Board for Professional Teaching Standards, which developed standards for accomplished teaching in more than 30 teaching areas defined by subject matter and developmental level of students. The Board then developed an assessment of accomplished teaching 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. These pieces of evidence are scored by trained raters who are expert in the same teaching field, using rubrics that define critical dimensions of teaching as the basis of the evaluation. Designed to identify experienced accomplished teachers, a number of states and districts, including the ones noted earlier, use National Board Certification as the basis for salary bonuses or other forms of teacher recognition, such as selection as a mentor or lead teacher. 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 more fairly to students who need them.

A number of recent studies have found that the National Board Certification assessment process identifies teachers who are more effective in raising student achievement than others who have not achieved certification.[2] Perhaps equally important, many studies have found that teachers' participation in the National Board process supports their professional learning and stimulates changes in their practice. Teachers note that the process of analyzing their own and their students' work in light of standards enhances their abilities to assess student learning and to evaluate the effects of their own actions, while causing them to adopt new practices that are called for in the standards and assessments.[3] Teachers report significant improvements in their performance in each area assessed -- planning, designing, and delivering instruction, managing the classroom, diagnosing and evaluating student learning, using subject matter knowledge, and participating in a learning community -- and observational studies have documented that these changes do indeed occur.[4]

National Board participants often say that they have learned more about teaching from their participation in the assessments than they have learned from any other previous professional development experience.[5] 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.[6]

Following on the work of the National Board, a consortium of more than 30 states, working under the auspices of CCSSO, created the INTASC standards for beginning teacher licensing. Most states have now adopted these into their licensing systems. In some states, teacher performance assessments for new teachers, modeled after the National Board assessments, are being used either in teacher education, as a basis for the initial licensing recommendation (CA, OR), or in the teacher induction period, as a basis for moving from a probationary to a professional license (CT).

These assessments require teachers to document their plans and teaching for a unit of instruction, videotape and critique lessons, and collect and evaluate evidence of student learning. Like the National Board assessments, beginning teachers' ratings on the Connecticut BEST assessment have been found to significantly predict their students' value-added achievement on state tests.[7] This finding is especially significant since the lowest-scoring candidates who do not pass the assessment are not allowed to gain a professional license or gain tenure in Connecticut, so the analysis had to deal with a truncated range that did not include most of those teachers. (Those who do not pass have the opportunity to attempt the assessment, but must pass by their 3rd year in teaching to remain in the profession.) About 10% of candidates in Connecticut do not pass the assessment. A study of predictive validity is currently underway for the Performance Assessment for California Teachers (PACT).

These assessments have also been found to help teachers improve their practice. Connecticut's process of implementing INTASC-based portfolios for beginning teacher licensing involves virtually all educators in the state in the assessment process, either as beginning teachers taking the assessment or as school-based mentors who work with beginners, as assessors who are trained to score the portfolios, or as expert teachers who convene regional support seminars to help candidates learn about the standards. Educators throughout the system develop similar knowledge about teaching and learn how principles of good instruction are applied in classrooms. These processes can have far-reaching effects. By the year 2010, an estimated 80% of elementary teachers, and nearly as many secondary teachers, will have participated in the new assessment system as candidates, support providers, or assessors.[8]

A beginning teacher who participated in the assessment described the power of the process, which requires planning and teaching a unit, and reflecting daily on the day's lesson to consider how it 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 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."

The same learning effects are recorded in research on the similar PACT assessment used in California teacher education programs. The assessment requires student teachers or interns to plan and teach a week-long unit of instruction mapped to the state standards; to reflect daily on the lesson they've just taught and revise plans for the next day; to analyze and provide commentaries of

videotapes of themselves teaching; to collect and analyze evidence of student learning; to reflect on what worked, what didn't and why; and to project what they would do differently in a future set of lessons. Candidates must show how they take into account students' prior knowledge and experiences in their planning. Adaptations for English language learners and for special needs students must be incorporated into plans and instruction. Analyses of student outcomes are part of the evaluation of teaching.

Faculty and supervisors score these portfolios using standardized rubrics in moderated sessions following training, with an audit procedure to calibrate standards. Faculties use the PACT results to revise their curriculum. In addition, both the novice teachers and the scoring participants describe benefits for teacher education and for learning to teach from the assessment and scoring processes. For example:

> For me the most valuable thing was the sequencing of the lessons, teaching the lesson, and evaluating what the kids were getting, what the kids 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 that you have together, but knowing that that's flexible and that it has to be flexible, based on what the children learn that day.
>
> -- Prospective teacher

> 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.
>
> -- 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.
>
> -- 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.
>
> -- Induction program coordinator

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. The TEACH Act contains a provision to develop a nationally available beginning teacher performance assessment, based on these models, which could provide a useful measure of effectiveness for new teachers and could leverage stronger accountability and improvement in teacher education.

Standards-Based Evaluations of Teaching. Similarly, standards-based teacher evaluations used by some districts have been found to be significantly related to student achievement gains for teachers and to help teachers improve their practice and effectiveness.[9] Like the teacher performance assessments described above, these systems for observing teachers' classroom practice are based on professional teaching standards grounded in research on teaching and learning. They use systematic observation protocols to examine teaching along a number of dimensions. All of the career ladder plans noted earlier use such evaluations as part of their systems and many use the same or similar rubrics for observing teaching. The Denver compensation system, which uses such an evaluation system as one of its components, describes the features of its system as including: well-developed rubrics articulating different levels of teacher performance; inter-rater reliability; a fall-to-spring evaluation cycle; and a peer and self-evaluation component.

In a study of three districts using standards-based evaluation systems, researchers found positive correlations between teachers' ratings and their students' gain scores on standardized tests (Milanowski, Kimball, & White, 2004). In the schools and districts studied, assessments of teachers are based on well-articulated standards of practice evaluated through evidence including observations of teaching along with teacher interviews and, sometimes, artifacts such as lesson plans, assignments, and samples of student work.

The Teacher Advancement Program offers one well-developed example of a highly-structured teacher evaluation system that was developed based on the standards of the National Board and INTASC and the assessment rubrics developed in Connecticut and Rochester (NY), among others.[10] In the TAP system of "instructionally-focused accountability," each teacher is evaluated four to six times a year by master / mentor teachers or principals who are trained and certified evaluators using a system that examines designing and planning instruction, the learning environment, classroom instruction, and teacher responsibilities. The training is a rigorous four-day process, and trainers must be certified based on their ability to evaluate teaching accurately and reliably. Teachers also study the rubric and its implications for teaching and learning, look at and evaluate videotaped teaching episodes using the rubric, and engage in practice evaluations. After each observation, the evaluator and teacher meet to discuss the findings and to make a plan for ongoing growth. Like other well-developed career ladder systems, TAP provides ongoing professional development, mentoring, and classroom support to help teachers meet these standards. Teachers in TAP schools report that this system, along with the intensive professional development offered, is substantially responsible for improvements in their practice and the gains in student achievement that have occurred in many TAP schools.[11] As described later, data from this extensive teacher evaluation and development system is combined with evidence about school-wide and individual teacher student achievement gains in making judgments about teachers' appointment to specific roles in the career ladder.

The set of studies on standards-based teacher evaluation suggest that the more teachers' classroom activities and behaviors are enabled to reflect professional standards of practice, the more effective they are in supporting student learning – a

finding that would appear to suggest the desirability of focusing on such professional standards in the preparation, professional development, and evaluation of teachers. These kinds of results led Hassell (2002) to conclude in his review of teacher pay systems that tying teachers' advancement and compensation to their knowledge and skills and using evaluation systems that help develop those skills, as these systems do, may ultimately produce more positive change in practice than evaluating teachers based primarily on student test scores.

Standards-based evaluation systems have also been used to evaluate beginning teachers for continuation and tenure and to identify struggling teachers for additional assistance and potential dismissal. The most long-standing evaluation systems that have successfully supported evaluation and personnel actions for both beginning and veteran teachers are those that have used Peer Assistance and Review Programs that rely on highly expert mentor teachers to conduct evaluations and provide assistance to teachers who need it. The systems in Rochester, New York; Cincinnati, Columbus, and Toledo, Ohio; and Seattle, Washington have all been studied and found successful in identifying teachers for continuation and tenure as well as intensive assistance and personnel action (see, e.g. NCTAF, 1996).

Key features of these systems include not only the instruments used for evaluation but also the expertise of the evaluators – skilled teachers in the same subject areas and school levels who have released time to serve as mentors to support their fellow teachers – and the system of due process and review that involve a panel of both teachers and administrators in making recommendations about personnel decisions based on the evidence presented to them from the evaluations.

In these systems, beginning teachers have been found to stay in teaching at higher rates because of the mentoring they receive, and those who leave (generally under 5%) are usually those the district has chosen not to continue rather than those who have quit. Among veteran teachers identified for assistance and review (usually 1-3% of the teaching force), generally about half improve sufficiently with intensive mentoring to be removed from intervention status and about half leave by choice or by district request. Because teacher associations have been closely involved in designing and administering these programs in collaboration with the district, the union does not bring grievances when a teacher is discontinued.

Evidence about Teachers' Knowledge, Skills, and Practices

For a variety of reasons, it can be important to document and reward in a teacher evaluation and compensation system aspects of teachers' knowledge and skills -- as well as their practices – that are associated with student learning. Schools need a mix of knowledge, skills, and abilities among their faculties to inform curriculum decisions and to meet the needs of their students. For example, aside from the knowledge of content and pedagogy teachers generally acquire in their certification area, specialized knowledge about the teaching of English language learners or the teaching of special education students may be highly desirable in many school contexts. Knowledge of the home languages students speak is also essential for communicating with parents as well as students. Proficiency in using

specific educational techniques, such as Reading Recovery or Cognitively Guided Instruction in mathematics, may be important in certain contexts.

The two-fold rationale for knowledge and skills-based compensation is that there should be incentives for teachers to continue to develop their abilities in ways that are important for student success, and there should be encouragement for teachers to use practices that have been found to be effective. As schools seek to offer a more coherent approach to instruction, encouragement for shared practices among teachers is also important. The kinds of knowledge, skills, and practices to be documented and recognized should be those known to be associated with greater individual and organizational effectiveness. As Odden and colleagues note:

> Knowledge- and skills-based compensation systems provide a mechanism to link pay to the knowledge and skills (and by extension, performance) desired of teachers....The concept of knowledge- and skills-based pay in education was adapted from the private sector, where it was developed to encourage workers to acquire new, more complex, or employer-specific skills. Knowledge- and skills-based pay was also intended to reinforce an organizational culture that values employee growth and development and to create a clear career path linked to increasing professional competence.[12]

Evidence that particular kinds of knowledge and skills impact student achievement can guide decisions about what should be documented and recognized. For example, there is evidence that a masters degrees in the field to be taught (e.g. mathematics or mathematics education) is associated with greater effectiveness,[13] as is training in how to work with diverse student populations (training in cultural diversity, teaching limited English proficient students, and teaching students with special needs).[14] In addition, some specific practices, such as the use of formative assessment to provide feedback to students and opportunities for them to revise their work, have been found in many dozens of studies to have large effect sizes on student learning gains.[15] Teachers who teach students specific meta-cognitive strategies for reading, writing, and mathematical problem solving have been found produce increased student learning of complex skills.[16] And so on.

In some systems, teachers receive recognition for demonstrating that they have implemented particular new practices like these associated with school-wide or district-wide goals, such as the use of common literacy practices across classrooms, or the use of formative assessments in planning and modifying instruction, or the implementation of a new system of writing instruction. Where possible, these practices are documented along with evidence of how the changes have affected student participation and learning. The rationale for using these measures of effective teaching practices is that they support teacher development and school-wide change initiatives, and are related to improvements in the conditions for student learning.

Odden and colleagues offer several examples of knowledge- and skills-based evaluation and compensation plans.[17] For example, Coventry, Rhode Island provides stipends for National Board Certification and for teachers to develop their skills in authentic pedagogy, self-reflection, differentiated instruction, and family and community involvement – all of which are strategies that have been linked

through research to student achievement. Douglas County, Colorado offers compensation for completing blocks of courses associated with district-goals, such as assessment or teaching diverse learners. Vaughan Learning Center, a charter school in Los Angeles, California, offers compensation for relevant degrees and certification, as well as for specific knowledge and skills relevant to the school's mission, such as literacy training, training for teaching English as a second language, special education inclusion, and technology.

Teacher proficiencies can be documented through systematic collection of evidence about planning and instruction, work with parents and students, and contributions to the school. This can be accomplished both through observations of practice, documentation of training or proficiencies, and a portfolio of teacher evidence about practices both in and beyond the classroom. In addition to specific teaching practices, a teacher might document how she increased student attendance or homework completion through regular parent conferences and calls home and show evidence of changes in these student outcomes, as well as other outcomes associated with them, such as improved grades, graduation, and college-going. Odden and colleagues note that a teacher portfolio in such a system "may include artifacts such as scholarly papers in the content area written by the teacher, new curricular the teacher has developed, logs of parental involvement, samples of tests and assignments, lesson plans, and essays reflecting on the teacher's practice."[18]

Evidence of Student Learning

Interest in including evidence of student learning in evaluations of teachers has been growing. After all, if student learning is the primary goal of teaching, it appears straightforward that it ought to be taken into account in determining a teachers' competence. At the same time, the literature includes many cautions about the problems of basing teacher evaluations substantially on student test scores. In addition to the fact that curriculum-specific tests that would allow gain score analyses are not typically available in many teaching areas, these include concerns about overemphasis on teaching to the test at the expense of other kinds of learning; problems of attributing student gains to specific teachers; and disincentives for teachers to serve high-need students, for example, those who do not yet speak English and those have special education needs (and whose test scores therefore may not accurately reflect their learning). This could inadvertently reinforce current practices in which inexperienced teachers are disproportionately assigned to the neediest students or schools discourage high-need students from entering or staying. At the same time, some innovative career ladder and compensation programs (in Rochester, New York and Denver, Colorado, for example, as well as the TAP system described earlier) have found valid ways to include evidence of student learning in teacher evaluations. These are discussed below.

The Use of Value-Added Achievement Test Scores to Evaluate Teachers. Because of a desire to recognize and reward teachers' contributions to student learning, a prominent proposal is to use *value-added student achievement test scores* from state or district standardized tests as a key measure of teachers' effectiveness. The value-added concept is important, as it reflects a desire to

acknowledge teachers' contributions to students' progress, taking into account where students begin. Furthermore, value-added methods are proving valuable for research on the effectiveness of specific populations teachers (for example, those who are National Board Certified or those who have had particular preparation or professional development experiences) and on the outcomes of various curriculum and teaching interventions.

However, there are serious technical and educational challenges associated with using this approach to make strong inferences about *individual* teacher effectiveness, especially for high-stakes purposes, as opposed to studying the effectiveness of *groups* of teachers in a research context. Among other things, for example, when researchers are aggregating data about large groups of teachers for research rather than decision-making purposes, they make various assumptions about how to treat missing student data, which students to include, or how to choose among models using different statistical controls that change the results of their estimates. Researchers may be concerned from an intellectual perspective about whether their models are indeed capturing teacher effects (as opposed to student variables or testing artifacts or the results of school practices outside the classroom), but they need not worry about whether their decisions disadvantage particular teachers in the way they would need to if these analyses were to be used to make individual personnel decisions.

Indeed, the emergent strategies being used to analyze student learning data to assess potential teacher effectiveness produce very different results depending on the different decisions researchers make about how to handle the data (for example, whether or not to control for student demographic characteristics or school effects, whether and how to interpolate missing data for students, whether to include or exclude special needs learners or new English language learners, whether to use tests that do not measure the specific curriculum a teacher teaches). Leading researchers agree that, while it is useful for research purposes, value-added modeling (VAM) is not appropriate as a primary measure for evaluating individual teachers. Summarizing the results of many studies, including a recent wide-ranging review by the RAND Corporation, Henry Braun of the Educational Testing Service concluded:

> VAM results should not serve as the sole or principal basis for making consequential decisions about teachers. There are many pitfalls to making causal attributions of teacher effectiveness on the basis of the kinds of data available from typical school districts. We still lack sufficient understanding of how seriously the different technical problems threaten the validity of such interpretations.[19]

The career ladder or compensation systems that do use student achievement data include it only as component of a broader system that incorporates evidence from standards-based evaluation systems, teacher performance assessments, or other evidence about teacher qualifications and practices. Often these data come from classroom, school, or district assessments rather than state tests, for reasons discussed further below. These data are

triangulated and interpreted to understand a teachers' practice in a multi-faceted way, rather than using a single measure to draw inferences that may be problematic.

The problems researchers have identified with using value-added testing models as a primary determinant of teacher effectiveness, especially those drawing on once-a-year large-scale assessments, include the following:

Teachers' ratings are affected by differences in the students who are assigned to them. Students are not randomly assigned to teachers – and statistical models cannot fully adjust for the fact that some teachers will have a disproportionate number of students who may be exceptionally difficult to teach (students with poor attendance, who are homeless, who have severe problems at home, etc.) and whose scores on traditional tests are problematic to interpret (e.g. those who have special education needs or who are English language learners). This can create both misestimates of teachers' effectiveness and disincentives for them to want to teach the students who have the greatest needs.

VAM requires scaled tests, which most states don't use. Furthermore, many experts think such tests are less useful than tests that are designed to measure specific curriculum goals. In order to be scaled, tests must evaluate content that is measured along a continuum from year to year. This reduces their ability to measure the breadth of curriculum content in a particular course or grade level. As a result, most states have been moving away from scaled tests and toward tests that measure standards based on specific curriculum content, such as end-of-course tests in high school that evaluate standards more comprehensively (e.g. separate tests in algebra, geometry, algebra 2, and in biology, chemistry, and physics). These curriculum-based tests are more useful for evaluating instruction and guiding teaching, but do not allow value-added modeling. Entire state systems of assessment that have been developed over many years – such as the New York State Regents system and systems in states like California, Washington, Massachusetts, Maine, Connecticut, Kentucky, and many more -- would have to be dismantled to institute value-added modeling.

VAM models do not produce stable ratings of teachers. Teachers look very different in their measured effectiveness when different statistical methods are used. Different teachers appear effective depending on whether student characteristics are controlled, whether school effects are controlled, and what kinds of students teachers teach (for example, the proportion of special education students or English language learners). In addition, a given teacher may appear to have differential effectiveness from class to class and from year to year, depending on these things and others. Braun notes that ratings are most unstable at the upper and lower ends of the scale, where many would like to use them to determine high or low levels of effectiveness.

Most teachers and many students are not covered by relevant tests. Scaled annual tests with previous year test results are not available in most states for teachers of science, social studies, foreign language, music, art, physical education, special education, vocational / technical education, and other electives in any grades, or for teachers in grades k-3 and nearly all teachers in grades 9-12. Furthermore,

because the scores are unstable, experts recommend at least 3 years of data for a given teacher to smooth out the variability. With many grades and subjects uncovered by scaled tests, and with three years of data needed to get a reasonably stable estimate for a teacher (thus excluding 1st and 2nd year teachers), at best only about 30% of elementary teachers and 10% of high school teachers would be covered by data bases in most states.

Missing data threatens the validity of results for individual teachers. Once teacher and student mobility are factored in, the number of teachers who can be followed in these models is reduced further. In low-income communities, especially, student mobility rates are often extremely high, with a minority of students stable from one year to the next. Although researchers can make assumptions about score values for missing student data for research purposes, these kinds of adjustments are not appropriate for the purposes of making individual teacher judgments.

Many desired learning outcomes are not covered by the tests that are widely used. Tests in the United States are generally much narrower than assessments used in other high-achieving countries (which feature a much wider variety of more ambitious written, oral, and applied tasks), and scaled tests are narrower than some other kinds of tests. For good or for ill, research finds that high-stakes tests drive the curriculum to a substantial degree. Thus, it is important that measures used to evaluate teacher effectiveness find ways to include the broad range of outcomes valued in schools. Otherwise, teachers will have little incentive to continue to include untested areas such as writing, research, science investigations, social studies, and the arts, or skills such as data collection, analysis, and synthesis, or complex problem solving, which are generally untested.

It is impossible to fully separate out the influences of students' other teachers, as well as school conditions, on their apparent learning. Prior teachers have lasting effects, for good or ill, on students' later learning, and current teachers also interact to produce students' knowledge and skills. For example, the essay writing a student learns through his history teacher may be credited to his English teacher, even if she assigns no writing; the math he learns in his physics class may be credited to his math teacher. Specific skills and topics taught in one year may not be tested until later years. A teacher who works in a well-resourced school with specialist supports may appear to be more effective than one whose students don't receive these supports. A teacher who teachers large classes without adequate textbooks or materials may appear to be less effective than one who has a small class size and plentiful supplies. As Braun notes, "it is always possible to produce estimates of what the model designates as teacher effects. These estimates, however, capture the contributions of a number of factors, those due to teachers being only one of them. So treating estimated teacher effects as accurate indicators of teacher effectiveness is problematic." To understand the influences on student learning, more data about teachers' practices and context are needed.

Thus, while value-added models are useful for looking at groups of teachers for research purposes – for example, to examine the results of preparation or

professional development programs or to look at student progress at the school or district level – and they may provide one measure of teacher effectiveness among several, they are problematic as the primary or sole measure for making evaluation decisions for individual teachers. In the few systems where such measures are used for personnel decisions such as performance pay, they are often used for the entire group of teachers in a school, rather than for individuals. Where they are used, they need to be accompanied by an analysis of the teachers' students and teaching context, and an evaluation of the teachers' practices.

Using Other Evidence of Student Learning. The fact that value-added analysis of test score data in large-scale testing systems is not always appropriate or available as a tool for evaluating individual teachers does not mean that states or districts cannot recognize and reward excellent teachers who produce strong student learning, or create incentives for them to help other teachers and serve the neediest students. It is possible to use other measures of student learning in evaluations of teaching, sometimes pre- and post-tests of learning conducted by districts or schools, or even learning evidence that is assembled by the teacher him or herself. Such evidence can be drawn from classroom assessments and documentation, including pre- and post-test measures of student learning in specific courses or curriculum areas, evidence of student accomplishments in relation to teaching activities, and analysis of standardized test results, where appropriate. The evidence can be assembled in a teaching portfolio by the teacher, demonstrating and explaining the progress of students on a wide range of learning outcomes in ways that take students' starting points and characteristics into account.

In some schools, teachers use their own fall and spring classroom assessments (or pre- and post-unit assessments) as a way of gauging student progress. These measures can also be tailored for the learning goals of specific students (for example, special education students or English language learners.) As part of a portfolio of evidence, these measures can document teacher effectiveness in achieving specific curriculum goals. Measures of student learning in specific subject areas may be scored writing samples or reading samples, mathematics assessments, assessments of science or history knowledge, or even musical performances. These typically provide better measures of classroom learning in a specific course or subject area because they are curriculum-specific and can offer more authentic measures of student learning. They are also more likely to capture the effects of a particular teacher's instruction and be available for most or all students. A teacher might even document the Westinghouse science competition awards she helped students win, or specific break-throughs achieved by her special education students, with evidence of her role in supporting these accomplishments.

In Denver's *Procomp* system,[20] for example, teachers set two goals annually in collaboration with the principal, and document student progress toward these goals using district, school, or teacher-made assessments to show growth. In Rochester's career ladder, evidence of student learning, determined by the teacher, is assembled in the teachers' portfolio. Arizona's career ladder program – which encourages local districts to design their own systems – requires the use of various methods of student assessment to ascertain teachers' effectiveness.

One study of the Arizona career ladder programs found that, over time, participating teachers demonstrated an increased ability to create locally-developed assessment tools to assess student learning gains in their classrooms; to develop and evaluate pre- and post-tests; to define measurable outcomes in "hard to quantify areas" like art, music, and physical education; and to monitor student learning growth in their action plans. They also showed a greater awareness of the importance of sound curriculum development, more alignment of curriculum with district objectives, and increased focus on higher quality content, skills, and instructional strategies.[21] Thus, the development and use of student learning evidence seemed to be associated with improvements in practice. In all of these career ladder systems, evidence of student learning is combined with evidence from standards-based teaching evaluations conducted through classroom observation, and evidence of teachers' skills or practices, as described below.

Implications for Policy

Efforts to recognize teacher competence and effectiveness as the basis for personnel decisions are not new in the policy arena, but recent initiatives have provided some potential break-throughs. Efforts to institute versions of merit pay or career ladders in the U.S. have faltered many times before – in the 1920s, the 1950s, and most recently in the 1980s, when 47 states introduced versions of merit pay or career ladders, all of which had failed by the early 1990s.[22] The reasons for failure have included faulty evaluation systems, concerns about bias and discrimination, pitfalls of strategies that rewarded individual teachers while undermining collaborative organizational efforts, dysfunctional incentives that caused unintended negative side-effects for serving all children, and lack of public will to continue increased compensation.

The initiatives detailed in this paper demonstrate that systems can provide recognition for demonstrated knowledge, skill, and expertise that move the mission of the school forward and reward excellent teachers for continuing to teach, without abandoning many of the important objectives of the current salary schedule -- equitable treatment, incentives for further learning, and objective means for determining pay. Promising beginnings have been made in some states and local districts that have developed new approaches to examining teacher performance and building career ladders. These approaches use multiple measures of performance, typically considering three kinds of evidence in combination with one another:

(1) Teachers' *performance on teaching assessments* measuring standards known to be associated with student learning (including national assessments, such as National Board Certification, and locally-managed standards-based teacher evaluations);

(2) Evaluation of *teaching practices* that are associated with desired student outcomes and achievement of school goals, through systematic collection of evidence about teacher planning and instruction, work with parents and students, and school contributions; and

(3) Contributions to *growth in student learning* (from classroom assessments and documentation as well as standardized tests, when appropriate).

All three of these strategies are used in the Denver, CO Procomp system of teacher compensation based on knowledge, skills, and performance; Rochester's Career in Teaching program; and Minnesota's Alternative Professional Pay System,[22] which were developed in collaboration with local or state teachers associations. Beyond recognizing teachers with new roles or compensation, these systems demonstrate that rewarding teachers for deep knowledge of subjects, additional knowledge in meeting special kinds of student and school needs, and high levels of performance measured against professional teaching standards can encourage teachers to continue to learn needed skills and enhance the expertise available within schools.

State and Local Initiatives

The work that has been done over the last decade to develop and assess teaching standards and to build new models of evaluation and recognition in school districts holds promise for creating more systematic means for developing teacher and teaching quality. Policies for identifying and supporting teacher and teaching effectiveness can be considered for both the beginning of the teaching career -- for licensing, hiring, and tenure decisions -- and for later stages of teacher development – for compensation and advancement decisions.

Identifying and Developing Beginning Teacher Effectiveness. It is important to be able to make licensing decisions based on greater evidence of teacher competence than merely completing a set of courses or surviving a certain length of time in the classroom. Since the 1980s, the desire for greater confidence in licensing decisions has led to the introduction of teacher licensing tests in nearly all states. However, these tests – generally multiple-choice tests of basic skills and subject matter – are not strongly predictive of teachers' abilities to effectively teach children. 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. Even paper-and-pencil tests of teaching knowledge, used in a few states, provide little evidence of what teachers can actually *do* in the classroom.

In the coming years, states will be able to benefit from the development of teaching performance assessments that evaluate teachers' practices related to student learning and have been found to be predictive of teachers' effectiveness. States now have the possibility of beginning to examine teacher performance as a basis for granting the initial probationary or later professional license, building on the work that has been done by some states and universities to build reliable and valid assessments that predict teacher effectiveness. Their work demonstrates that on-the-job performance assessments of beginning teachers can be used during teacher education (at the end of an internship or student teaching) as the basis for a licensure recommendation. Systematically scored portfolios including direct evidence of teaching have been developed with state encouragement or requirement by universities in Vermont, Maine, Wisconsin, Oregon, and California. Oregon's teacher Work Sampling System provides pre- and post-test evidence of teachers' contributions to student learning, constructed by teachers themselves.

California's teacher performance assessment, described earlier, which also includes evidence of student learning in relation to a unit of teaching, will be a funded, statewide requirement by 2008.

Some states have also used performance assessments of first or second year teachers (during their probationary period) as the basis for granting a professional license (usually acquired in the 3rd year of practice) and, by implication, setting a clear bar for the tenure decision. Connecticut's system is most highly developed and reliably scored, but initiatives have also been undertaken in North Carolina and California as part of state induction programs.

All of these initiatives have been based on the beginning teacher licensing standards developed by the Interstate New Teacher Assessment and Support Consortium (INTASC), sponsored by the Council of Chief State School Officers. An effort by this consortium to fine-tune and pilot this work more broadly could give momentum to an effort to better evaluate teacher competence and effectiveness at the beginning of the teaching career.

States can also encourage and support localities in developing stronger evaluation of beginning teachers in the early years prior to tenure, tied to effective mentoring from highly accomplished veterans that will help novices meet the standards. Most states now require an induction program of some sort and many also provide some level of funding. However, the activities that are to occur during the induction process and the type of teaching to be developed are often not specified, so programs are frequently less powerful than they could be.

Connecticut wraps its required mentoring of beginning teachers around the teacher performance assessment so that the standards of performance are clear. High-quality local standards-based evaluations, like those described earlier, can also be used for this purpose. Organizing mentoring around clear standards of practice that have been tied to teacher effectiveness focuses the mentor's and novice's efforts on what matters most for teaching success. Of course, this strategy also requires highly-skilled mentors who are themselves effective teachers. This leads to the question of how to identify and select such leaders.

Identifying and Developing Teacher Effectiveness Throughout the Career. If teachers are better supported and selected for tenure in the early years of the career, the prospects for developing a highly effective teacher corps will be much enhanced. As we have noted, progress has been made in developing career development systems that can recognize excellent teaching and both reward it and tap the knowledge of such teachers on behalf of broader school improvements. These initiatives generally have several features in common. All require teacher participation and buy-in to be implemented. Typically, evaluations occur at several junctures as teachers move from their *initial license*, through a period as a *novice or resident teacher* under the supervision of a mentor, to designation as *professional teacher* after successfully passing an assessment of teaching skills. *Tenure* is a major step tied to a serious decision made after rigorous evaluation of performance in the first several years of teaching, incorporating administrator and peer review by expert colleagues. *Lead teacher status* – which triggers additional compensation and access to differentiated roles -- may be determined by advanced certification from the National Board for Professional Teaching Standards and other evidence of

performance through standards-based evaluation systems. Such systems both encourage and measure effective teaching, and can be combined with other evidence of desirable teacher practices and student learning to identify accomplished teachers.

Where this has been done, it has proved critically important to design evaluation systems that provide a comprehensive picture of what teachers do and with what results, to be sure that evaluations are conducted reliably and validly by skilled assessors, and to be confident that evidence about student learning is carefully interpreted and properly attributed to the teacher.

Beyond the features of the evaluation systems, there are important lessons about the features of the policy systems in which they operate. For example, the system should be designed to operate so that teachers are not penalized for teaching the students who have the greatest educational needs. This requires sensitivity to student and classroom characteristics in the evaluation system. Furthermore, incentives should operate to support collegiality by recognizing all the teachers who reach specific criteria, rather than pitting teachers against each other in a situation in which one teacher's gain is another's loss.

The challenges to be overcome in designing productive systems for recognizing and rewarding teacher effectiveness were vividly illustrated by the testimony of an expert veteran teacher in Springfield, Massachusetts last year – a district being asked to put in place a system of merit pay based on value-added student achievement test scores. Springfield is a severely under-resourced district serving a predominantly minority, low-income student population. Fiscal woes had prevented salary increases for three years, and about half of the 2600 teachers in the district had left over this time. Nearly 25% of the teaching force was uncertified and inexperienced.

Susan Saunders, a Springfield native with more than 20 years of experience, was one of the local heroes who had stayed and worked tirelessly to assist the revolving door of beginning teachers, who shared the few updated textbooks with these teachers, and who took on the highest need special education students (comprising more than half of her class of 32 students). When asked how she would feel about working in this new system of test-based merit pay, Saunders said the introduction of the system would force a teacher like herself either to leave or change her approach entirely – to keep the best materials for herself, stop taking on the special education students, and stop helping the other teachers in her building (since one teacher's greater success would come at the expense of another teacher's rating).

The Springfield system was not adopted because an arbitrator deemed the technical validity of the proposed system inadequate to carry the weight of personnel decision making. This example suggests how important it is to exercise care in developing systems of rewards for teachers so they do not create incentives that would discourage teachers from working collaboratively with each other and taking on the most challenging students. Since any measures used are likely to drive instruction, it is also critically important that the assessments used to evaluate student learning cover the broad goals of learning that are valued and are valid for the students whose results would be considered.

State encouragements for local career ladders and innovative compensation systems, like those in Minnesota and Arizona, can be designed to ensure that several important features are in place. These would include:

(1) Teacher collaboration and buy-in in developing the system;

(2) Recognition and encouragement of collegial contributions to overall school success and clear criteria for accomplishment that all eligible teachers can achieve, rather than a quota system that pits teachers against each other;

(3) Valid evidence of teacher effectiveness based on multiple measures, including:

(3.1) standards-based evaluation of practice, such as National Board Certification, a valid state teacher performance assessment; or local evaluations of teacher performance;

(3.2) evidence of practice based on multiple classroom observations and examination of other classroom evidence (e.g. lesson plans, student assignments and work samples) by multiple evaluators using a standards-based evaluation instrument that examines planning, instruction, the learning environment, and student assessment.

(3.3) evidence of learning of the teacher's students on valid assessments that appropriately evaluate the curriculum the teacher teaches;

(4) Consideration of the needs of the students the teacher serves and valid and appropriate assessment of all students included in the analysis, including students with special learning needs and new English language learners,

(5) Ongoing, high-quality professional learning opportunities to enable teachers to learn to meet the standards.

Policy Possibilities

Given the challenges to be surmounted in designing and implementing new systems for identifying and recognizing teacher effectiveness, the role of policy should be supportive rather than directive. There are many things to be learned about how to measure teacher effectiveness in ways that are accurate and valid, that create knowledge and incentives for strong collegial work and for teaching all students well. Only a few dozen districts have been able to launch career ladders that have worked and lasted for more than a few years. Any effort to stimulate more productive work in this area should initially provide incentives to state and local initiatives that can garner support and develop models with potential for scale-up.

There are three areas where governmental support could be particularly helpful:

1) To develop and measure beginning teacher effectiveness, fund research and development to make available a beginning teacher performance assessment, along with support for beginning teacher mentoring. Initial teacher competence and effectiveness could be better ascertained, and preparation and mentoring could be strengthened, if they were guided by a high-quality, nationally-available teacher performance assessment, which measures actual teaching skill in the content areas, and which can guide teacher learning and help to develop sophisticated practice as part of licensing and ongoing career advancement.

In the U.S., the Interstate New Teacher Assessment and Support Consortium (INTASC), sponsored by the Council of Chief State School Officers, has already created teacher licensing standards adopted by most states and has piloted performance assessments tied to the standards; several states, including Connecticut and California, have incorporated such performance assessments in the licensing process. As proposed in the TEACH Act, federal support to a consortium of states in concert with appropriate professional associations could further refine and pilot these assessments to provide a useful tool for accountability and improvement that would also facilitate teacher mobility across states by supporting license reciprocity.

Ideally, such a tool would be accompanied by a federally-funded incentive to states and districts to create strong mentoring programs for all beginning teachers. A matching grant program could ensure support for every new teacher in the nation through investments in state and district mentoring programs. Based on the funding model used in California's Beginning Teacher Support and Assessment Program, for example, a federal allocation of $4000 for each beginning teacher, matched by states and/or local districts, could fund mentoring for every novice teacher (about 125,000 annually)[24] for an investment of $500 million a year. If even half of the early career teachers who currently leave teaching were to be retained, the nation would save at least $600 million a year in replacement costs while gaining more competent teachers.

2) Provide incentive funds for states and localities to develop systems that recognize and tap teacher expertise, and to reward accomplished teachers who take leadership roles in high-need schools. The federal government could encourage districts to develop systems that recognize effective teachers and create career ladders that tap their skills through a competitive grants program. To build teacher effectiveness, such initiatives would incorporate beginning teacher mentoring as well as stages in the career enabling a broader range of roles for expert teachers. They would be accompanied by performance-based teacher evaluation systems that provide information about teacher effectiveness through standards-based teacher evaluations well as systematic collection of evidence about teachers' practices and student learning. Such systems should include evidence of high-quality professional learning opportunities and school designs that provide time for teachers to work and learn together during the school day. They should also be designed to build collaborative incentives and to recognize and support teachers who teach the highest-need students.

A federal initiative could include additional incentives for the design of innovative approaches to attract and keep accomplished teachers in priority low-income schools, through compensation for accomplishment and for additional responsibilities, such as mentoring and coaching. For example, $500 million would provide $10,000 in additional compensation for 50,000 teachers annually, to be allocated to expert teachers in high-need schools through state- or locally-designed incentive systems. (Matched by state and local contributions, this program would provide incentives to attract 100,000 accomplished teachers to high-poverty schools.)

Teacher expertise could be recognized through such mechanisms as National Board Certification, state or local standards-based evaluations, and carefully

assembled evidence of contributions to student learning. Incentives might also be structured to encourage such highly effective teachers, as part of a group of teachers, to take on redesigning and reconstituting failing schools so that they become more effective.

3) Support research on value-added modeling and other means for examining student learning growth. Given the interest in using student learning data in evaluations of teachers, and the challenges of doing so, it would be productive for the federal government to fund an impartial group of experts, through the National Academy of Sciences or the National Academy of Education, to examine the data systems and methodologies needed to use student learning data appropriately in systems that assess teaching.

Conclusion

Initiatives to measure and recognize teacher effectiveness appear to be timely, as the press for improved student achievement is joined to an awareness of the importance of teachers in contributing to student learning. Such initiatives will have the greatest pay-off if they are embedded in systems that also develop greater teacher competence through mentoring and coaching around the standards and through roles for teachers to help their colleagues and their schools improve. Initiatives will have a greater likelihood of survival and success if they also build confidence in the validity of the measures and create incentives for teachers to work with colleagues and teach the neediest students. Federal, state, and local partnerships to create increasingly valid measures of teacher effectiveness and to support the development of innovative systems for recognizing and using expert teachers can make a substantial difference in the recruitment and retention of teachers to the places they are most needed and, ultimately, in the learning of students.

Endnotes

[1] For a summary of studies, see L. Darling-Hammond & J. Bransford, *Preparing Teachers for a Changing World: What Teachers should Learn and Be Able to Do.* San Francisco: Jossey-Bass, 2005; L. Darling-Hammond (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8*(1), http://epaa.asu.edu/epaa/v8n1; Wilson, S.M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations.* A research report prepared for the U.S. Department of Education. Seattle: Center for the Study of Teaching and Policy, University of Washington.

[2] See for example, Bond, L., Smith, T., Baker, W., & Hattie, J. (2000). *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); Cavaluzzo, L. (2004). *Is National Board Certification an effective signal of teacher quality?* (National Science Foundation No. REC-0107014). Alexandria, VA: The CNA Corporation; Goldhaber, D., & Anthony, E. (2005). *Can teacher quality be effectively assessed?* Seattle, WA: University of Washington and the Urban Institute; Smith, T., Gordon, B., Colby, S., & Wang, J. (2005). *An examination of the relationship of the depth of student learning and National Board certification status* (Office for Research on Teaching, Appalachian State University). Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National Board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46), 117.

[3] Steven Athanases (1994). Teachers' reports of the effects of preparing portfolios of literacy instruction. *Elementary School Journal, 94*(4), 421-439.

[4] Edward Chittenden, & J. Jones (1997, April). *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, Chicago, IL; Sato, M. (2000, April). *The National Board for Professional Teaching Standards: Teacher learning through the assessment process.* Paper presented at the Annual Meeting of American Educational Research Association. New Orleans, LA; Tracz, S.M., Sienty, S. & Mata, S. (1994, February). *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. Chicago, IL; Tracz, S.M., Sienty, S. Todorov, K., Snyder, J., Takashima, B., Pensabene, R., Olsen, B., Pauls, L., & Sork, J. (1995, April). *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. San Francisco, CA.

[5] Haynes, D. (1995). One teacher's experience with National Board assessment. *Educational Leadership, 52* (8): 58-60; Bradley, A. (1994, April 20). Pioneers in professionalism. *Education Week, 13,* 18-21; Areglado, N. (1999, Winter). I became convinced: How a certification program revitalized an educator. *National Staff Development Council,* 35-37; Buday, M., & Kelly, J. (1996). National Board certification and the teaching professions commitment to quality assurance. *Phi Delta Kappan, 78*(3), 215-219.

[6] Haynes, p. 60.

[6] Wilson, M. & Hallum, P.J. (2006). 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.

[7] Pecheone, R. & Stansbury, K. (1996). Connecting teacher assessment and school reform. *Elementary School Journal, 97*, 163-177 (p. 174).

[8] Milanowski, A.T., Kimball, S.M., White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement.* University of Wisconsin-Madison: Consortium for Policy Research in Education.

[9] The teacher responsibility rubrics were designed based on several teacher accountability systems currently in use, including the Rochester (New York) Career in Teaching Program, Douglas County (Colorado) *Teacher's Performance Pay Plan,* Vaughn Next Century Charter School (Los Angeles, CA) Performance Pay Plan, and Rolla (Missouri) School District Professional Based Teacher Evaluation.

[10] Lewis Solomon, J. Todd White, Donna Cohen & Deborah Woo (2007). *The effectiveness of the Teacher Advancement Program.* National Institute for Excellence in Teaching, 2007

[11] Allan Odden, Carolyn Kelley, Herbert Heneman, and Anthony Milanowski (2001, November). Enhancing teacher quality through knowledge- and skills-based pay. *CPRE Policy Briefs, R-34.* Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

[12] Goldhaber, D.D., & Brewer, D.J. (1997). Evaluating the effect of teacher degree level on educational performance. In W.J. Fowler (Ed.), *Developments in School Finance, 1996* (pp.197-210). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

[13] Wenglinsky, H. (2002). The link between teacher classroom practices and student academic

performance. *Education Policy Analysis Archives, 10*(12).

[14] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment and Education: Principles, policy and practice, 5*(1), 7-75.

[15] See Darling-Hammond & Bransford, 2005, for example.

[16] Odden et al., 2001.

[17] Odden et al., 2001, p. 4.

[18] Henry Braun, *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models* (Princeton, NJ: ETS, 2005), p. 17.

[19] For more detail about the Denver Procomp system, see http://denverprocomp.org.

[20] Richard Packard & Mary Dereshiwsky (1991). *Final quantitative assessment of the Arizona career ladder pilot-test project.* Flagstaff: Northern Arizona University.

[21] Darling-Hammond, L., and Berry, B. (1988). *The Evolution of Teacher Policy.* Santa Monica, CA: RAND Corporation, 1988.

[22] For more detail about Denver, see http://denverprocomp.org. For more detail about the Minnesota plan see http://www.educationminnesota.org/index.cfm?PAGE_ID-15003.

[23] About 250,000 teachers are hired each year, but typically only 40-60% of them are new to teaching. The others are experienced teachers changing schools or returning teachers who are re-entering the labor force.

About the Author

Linda Darling-Hammond is Charles E. Ducommun Professor of Education at Stanford University where she has launched the Stanford Center for Opportunity Policy in Education and the School Redesign Network and served as faculty sponsor for the Stanford Teacher Education Program. She is a former president of the American Educational Research Association and member of the National Academy of Education. Her research, teaching, and policy work focus on issues of school reform, teacher quality and educational equity. From 1994-2001, she served as executive director of the National Commission on Teaching and America's Future, a blue-ribbon panel whose 1996 report, What Matters Most: Teaching for America's Future, led to sweeping policy changes affecting teaching in the United States. In 2006, this report was named one of the most influential affecting U.S. education and Darling-Hammond was named one of the nation's ten most influential people affecting educational policy over the last decade. She recently served as the leader of President Barack Obama's education policy transition team.

Contact information: Stanford University School of Education, Stanford, CA 94305 ; telephone: 650-723-3555; email: ldh@stanford.edu

This article was reprinted with permission from the Council for Chief State School Officers.