Student Learning, Student Achievement:
HOW DO TEACHERS MEASURE UP?

A REPORT BY THE
Student Learning, Student Achievement Task Force

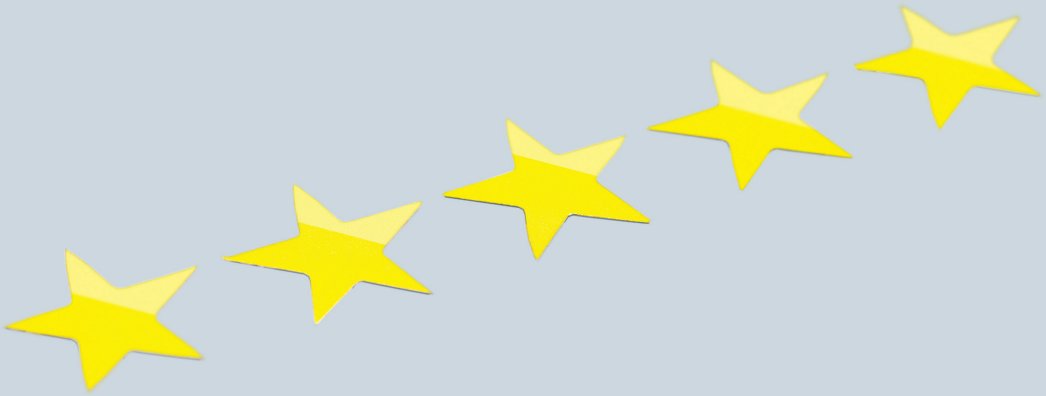**National Board for Professional Teaching Standards**

Student Learning, Student Achievement:

HOW DO TEACHERS MEASURE UP?

A REPORT BY THE

Student Learning, Student Achievement Task Force

National Board for Professional Teaching Standards

Student Learning, Student Achievement
TASK FORCE

CHAIR   Robert Linn, *NBPTS Certification Council Member*

MEMBERS   Lloyd Bond, *Professor Emeritus, University of North Carolina at Greensboro, and Senior Scholar Emeritus, the Carnegie Foundation*

Peggy Carr, *Associate Commissioner, Assessment Division, National Center for Education Statistics*

Linda Darling-Hammond, *Charles E. Ducommun Professor of Education, Stanford University*

Douglas Harris, *Associate Professor of Educational Policy Studies, University of Wisconsin at Madison*

Frederick Hess, *Resident Scholar and Director of Education Policy Studies, American Enterprise Institute for Public Policy Research*

Lee Shulman, *President Emeritus, Carnegie Foundation, and Charles E. Ducommun Professor Emeritus, Stanford University*

NBPTS STAFF

CHIEF PROGRAM OFFICER   Joan Auchter

EDITORIAL CONSULTANTS   Robert Johnston, *The Hatcher Group*

Mark Toner, *CommunicationWorks*

Lisa Towne

Acknowledgements

Many people contributed to this task force report. We are grateful to the National Board for Professional Teaching Standards for supporting the work of the task force. We thank Lisa Towne for her editorial support in the initial phases of the effort and Robert Johnston and Mark Toner for their editorial input through several drafts of the report. We are also grateful to Joan Auchter and her staff for the excellent support they provided the task force throughout the project. They worked tirelessly to facilitate the work of the task force and to bring the project to closure. Finally, as chair of the task force, I thank the task force members for their dedication and outstanding contributions to the report. They gave generously of their time. They not only shared their perspectives on issues addressed in the report, but they also listened to and respected the perspectives of other task force members.

Robert L. Linn, *Chair*
Student Learning, Student Achievement Task Force
National Board of Professional Teaching Standards

Table of Contents

Foreword From the
NBPTS President

The National Board for Professional Teaching Standards (NBPTS) welcomes the efforts of federal, state, and local policymakers to find new ways to ensure an accomplished teacher for every student in America. The National Board has advanced this mission since its inception in 1987. Today, that mission is carried out by the tens of thousands of National Board Certified Teachers (NBCTs) nationwide—each of whom completed the National Board's rigorous assessment process to demonstrate his or her competence in their teaching field.

Policymakers are right to want to link teacher evaluation to student performance as part of these efforts. Understanding how student learning and achievement can be measured and linked to the efforts of teachers has been of utmost importance to our work. We welcome initiatives that advance this understanding and translate new knowledge into ideas that can improve classroom teaching. Such advances have implications beyond individual NBCTs because we know that many of these teachers become mentors, teacher trainers, and school leaders. Improving how student performance is incorporated into teacher evaluation inevitably will influence practice at all of these levels.

At the same time, we must proceed carefully. As we have learned, such evaluations will be valid and relevant only if they are fair, accurate, and not limited to a single measure of teacher influence and effectiveness. If we do not get it right, the nation will lose a valuable opportunity to advance and improve teaching practice.

As a leader in teacher assessment and development, NBPTS is taking steps to ensure that the ongoing conversation about teacher evaluation will be rich, research-based, and reflective of various approaches. One lesson we have learned from years of refining how we evaluate accomplished teaching in 25 certification areas is that we must constantly reflect on our practices. That means asking some of the most thoughtful people in the field for their thinking, input, and even constructive criticism.

To further our understanding of how teachers are assessed in a new era of school improvement, NBPTS extended an invitation to several leaders in education evaluation, research, and policy. We asked them to participate in a series of conversations, share their collective knowledge, and then recommend how the National Board can strengthen its own work in this area while also continuing to be a leading source of information for the field.

....

If those of us involved in this effort do not get it right, the nation will lose a valuable opportunity to advance and improve educator evaluation.

....

The result of this important and thoughtful work is summarized in this white paper, which we are proud to share. This paper also includes several compelling recommendations that the National Board will consider in its future work. We look forward to drawing from this conversation and the resulting recommendations to steer National Board Certification and the field to better evaluation of accomplished teaching that builds an even stronger link to how our children learn and succeed in school.

Joseph A. Aguerrebere, Ed.D.

President and CEO
National Board for Professional Teaching Standards

Executive Summary
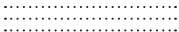
Advances in education data systems, measurement models, and practice-based research give us an opportunity to refine the meaning and identification of accomplished teaching. As a leader in identifying accomplished teaching, the National Board for Professional Teaching Standards (NBPTS) has convened a Student Learning, Student Achievement Task Force to study how it can continue playing a defining role in this new era. Made up of experts in assessment, school reform, and measuring teacher quality, the task force outlines in this white paper new methods of evaluating teachers' impact on student learning. Its recommendations are intended not only to improve the National Board Certification process, but also to provide guidance to the entire education community about appropriate ways to ground teacher evaluation in student learning.

Since its inception, the National Board's focus on the connection between accomplished teaching and student learning has been guided by a simple premise: the hallmark of accomplished teaching is student learning. NBPTS believes that the success of teachers in promoting student learning should be a defining measure of teacher quality. This simple but critical belief can be better realized because of the advances in applied assessment, technology, data systems, and test-based accountability models since the National Board's inception. Twenty years ago, the requisite systems did not yet exist, so any effort to identify accomplished teachers had to rely almost entirely on expert evaluations of teaching practice. Today, advances have made it increasingly possible to incorporate direct and systematic evidence about student learning into measurements of teacher quality.

Fulfilling this aspiration will include evaluating teachers on how well they help children learn across the breadth and depth of the curriculum. To meet this challenge, two issues must be addressed and were studied by the task force. The first issue is the tendency to rely primarily on achievement tests in a few grades and subjects to determine teacher effectiveness, to the exclusion of other subjects, grade levels, domains of learning, and evidence about teacher performance.

The other important factor pertains to the critical distinction between student learning and student achievement. Although the two terms are often used interchangeably, they convey profoundly different ideas, particularly as they relate to teaching. In brief, *student achievement* is the <u>status</u> of subject-matter knowledge, understanding, and skills at one point in time, while *student learning* is the <u>growth</u> in subject-matter knowledge, understanding, and skills over time. It is student learning—not student achievement—that is relevant to defining and assessing accomplished teaching.

In an attempt to measure student learning, many growth models have been developed. Of those models, the "value-added" approach has emerged as the method of choice to estimate the contributions that specific teachers and schools make to the growth in student learning. But while value-added models place necessary focus on important student outcomes, they remain constrained by technical issues involving the nature of tests, data quality, and the appropriate application of statistical models and methodologies. As we explain in greater detail later, even with better assessments, there will always be challenges in determining how much each teacher contributes to student learning. Education is a complex process with many actors, including teachers, principals, tutors, reading coaches, librarians, and—perhaps most important—parents. For this reason, thoughtful evaluations of teacher performance must combine direct evidence of student learning such as "value-added" data and examinations of teaching practice. Gains in student learning must always be examined within the context of teaching practice to ensure that they are connected to what teachers are doing in the classroom.

To better understand the complexities surrounding measurements of student learning and their role in the evaluation of teacher effectiveness, the Student Learning, Student Achievement Task

Force, which includes some of the National Board's most articulate critics, was charged with:

- Describing how student learning and achievement are captured in the National Board's evidence-based standards and certification process;

- Defining the critical distinction between student achievement and student learning;

- Identifying traditional and alternative approaches to measuring student learning; and

- Evaluating the strengths and limitations of these approaches as measures of teacher effectiveness.

Drawing on the National Board's quarter-century of certifying highly skilled teachers across all grade levels, more than 20 content areas, and all 50 states, the District of Columbia, and territories, the task force seeks to inform NBPTS and the broader education community of ways to effectively apply new tools, data systems, and technologies. Motivated by the belief that a teacher's contribution to student learning is the hallmark of accomplished teaching, the task force offers a series of principles and recommendations to guide the use of assessments of student learning as a measure of teacher effectiveness. Such measures should:

- **Be aligned with the curriculum and student learning goals a specific teacher is expected to teach**. Measures of student learning must reflect the specific content of what is expected to be taught. This principle also recognizes the importance of identifying the specific teacher or teachers responsible for gains in student learning, particularly given the fact that learning is a cumulative process, with previous teachers and learning experiences playing significant roles.

- **Be constructed to evaluate student learning**—that is, performance at two or more points in time—rather than a snapshot of student achievement, so that changes in student understanding and performance can be substantially attributed to instruction. This principle ap-

plies with equal force to standardized quantitative measures and more qualitative measures of student learning, such as portfolios of student work, both of which must focus on the students' gains in learning over the period a teacher provided instruction.

::::::: **Be sensitive to the diversity of students**, including those with special needs or limited English proficiency, as well as gifted and high-achieving students. Assessments used to evaluate teachers must be valid for the student populations they teach.

::::::: **Capture learning validly and reliably at the student's actual achievement level.** Measures should be evaluated continuously to determine the extent to which they address the principles of alignment with the range of knowledge and skills to be measured and the ability to capture student learning across the diverse learning needs and backgrounds outlined in this white paper.

::::::: **Provide evidence about student performance and teacher practice that reflects the full breadth of subject-matter knowledge and skills that are valued**. This recommendation addresses the need to identify the extent to which a teacher's practices are connected to and influence student learning. Linking these measures enables a rich and nuanced assessment of on-the-ground practice in context and can capture the complexities of the effects of teaching on student learning over time.

These principles are intended to serve as guidelines in designing teacher assessment systems that reflect student learning and improve teaching practice. We view the challenges in creating such systems as substantial—but not insurmountable, particularly if policymakers carefully evaluate the strengths and weaknesses of varying approaches to assessing student and teacher performance. To that end, the task force believes that National Board Certification should ultimately be a measure of how accomplished teachers are contributing to student learning. While the National Board Certification process already requires teachers to demonstrate multiple examples of student learning, we recommend that the NBPTS:

1   **Explore strengthening the extent to which student learning is systematically evaluated in each of the 25 certificate areas.** The task force recommends that the National Board strengthen evidence of student learning in each certification area and be more clear and precise about the nature of student work submitted in the portfolio process so that the work more accurately measures student learning in relation to teaching practice.

2   **Explore adding additional evidence of student learning, both created by teachers and from broader assessment measures, to the basket of evidence currently used in the National Board Certification process.** Following models such as those explored in this paper, NBPTS could, for example, develop criteria for using standardized assessment results in programs that tie teacher evaluation to student learning. It could also require teachers to submit on a pilot basis existing state or district assessment data, where aligned, valid, and available, as well as alternative measures of student learning in school districts and subject areas to augment standardized data or where such standardized data are not available.

3   **Continuously monitor research on the impact of teachers on student learning.** As the body of research continues to emerge, NBPTS should continually study the evidence and test the validity of its own standards and instruments against the evidence.

4   **Through the National Board's research, promote systematic use of methods for evaluating teachers' effectiveness and impact on student learning.** The National Board should conduct research and share the results with other stakeholders to help inform the use of information and assessments of both student learning and teacher effectiveness.

5 **Promote the development of teacher skills in designing classroom assessments and interpreting external assessment results, providing appropriate feedback to students, and using measures of student learning as a central element of accomplished teaching**. These are important aspects of teacher practice that bear directly on how much teachers contribute to student learning, and the more sophisticated teacher-created classroom assessments that would result from the development of such skills could become a strong component of the National Board Certification process.

The task force report underscores the need for educators and policymakers to combine smart measures of student learning with sensible efforts to identify accomplished teaching practice. Its members believe that by reflecting on its own efforts and constantly trying to refine and improve them, and by communicating to other stakeholders the broad principles guiding this effort and the insights that emerge, NBPTS will continue to play a leading role in identifying what both accomplished teachers and high-achieving students are expected to know and be able to do. ★

Setting the Stage

The core mission of the National Board for Professional Teaching Standards (N B P T S) is to create common standards for accomplished teaching and, through evidence-based assessments, credential accomplished teachers. Since its inception in 1987, the National Board has grappled with two crucial and complex questions: What assessment methods are most likely to credential accomplished teachers? And more to the point, how can measures of student learning and student achievement be used to measure accomplished teaching?

Nearly a quarter-century ago, the National Board's ambitious attempt to develop a large-scale, performance-based teacher assessment program was novel (National Research Council, 2008). However, with advances in measurement models, data systems, and practice-based research, the notion of linking teacher performance to student performance has gained prominence in the broader policy discourse around how to support, evaluate, reward, and retain high-quality teachers in the nation's schools. That notion is poised to become a focal point of a new era of school improvement. For example, states seeking federal stimulus dollars have been required to

············

The task force is guided by a simple premise: the hallmark of accomplished teaching is student learning.

············

provide the U.S. Department of Education information on "whether or not teachers are evaluated based on how well their students perform" and, more specifically, on "the number and percent of Local Educational Agencies teacher and principal evaluation systems that require evidence of student achievement outcomes." In addition, one of the highest priorities of the Bill & Melinda Gates Foundation's recent k–12 education initiative is to invest in districts that are willing to make teacher effectiveness the core of what they do in hiring, compensation, tenure, and placement decisions.

These changes offer a historic opportunity to refine the meaning and identification of accomplished teaching. It is in this context that the National Board convened the Student Learning, Student Achievement Task Force to take a fresh look at these questions. The task force's work seeks to help build on the National Board's efforts to link its assessments to student learning and to provide insight to others addressing similar issues.

Specifically, the National Board charged the task force with preparing a white paper that does the following:

- Defines how student learning and student achievement are captured in the NBPTS standards and certification process.

- Develops a working definition of the critical distinction between student learning, which measures growth in subject-matter knowledge, understandings, and skills over time, and student achievement, a subset of student learning that presents a snapshot of subject-matter knowledge, skills, and understanding at one point in time.

- Identifies traditional and alternative approaches to measuring student learning and achievement and the strengths and limitations of these approaches as measures of teacher performance effectiveness.

- Recommends to both the National Board and other stakeholders ways to improve the validity and reliability of student learning measures as a component of teacher evaluation.

The task force is guided by a simple premise: *the hallmark of accomplished teaching is student learning.* This intuitive yet powerful statement anchors teaching to its primary purpose—students becoming increasingly knowledgeable and skilled. For us, the question is not *whether* student learning ought to drive the certification of accomplished teaching, or any other teacher assessment for that matter. Rather, we are concerned with *how* student learning should drive that process in valid and practical ways.

Professional standards across a wide range of fields typically have been developed based on logic and professional consensus (and, to a lesser degree, research). Creating standards and evidence-based assessments that are closely related to outcomes—which is fundamentally what the current policy agenda involves—is inherently difficult. A National Academies committee that recently considered the impact of National Board Certification on student learning put it this way:

> Measures of outcomes for students, such as their academic achievement, do provide a means of evaluating teachers' job performance, but … it is enlightening to consider what this would mean if extrapolated to other fields. For example, this is similar to evaluating the validity of a medical certification test by collecting information about the outcomes for patients of a board-certified physician or evaluating the validity of the bar exam by considering the outcomes for clients of a lawyer who had passed the bar exam and been admitted to the bar. Outcomes for patients reflect many factors other than the skills and knowledge of the physician who provides services, such as the severity of the illness being treated and the degree to which the patient adheres to the professional advice given. Likewise in law, the outcome for the client depends on such factors as the nature of the legal problem, the record of prior legal problems, and the extent to which the client follows the [lawyer's] advice. Furthermore, should the outcomes for a high-priced lawyer, who can select his or her clients, be compared to the outcomes for a public defender? While data are available that might be used in such evaluations (e.g., rates of death or guilty verdicts) and several such studies have been conducted … many factors can contribute to the outcomes, making interpretation of the relationships very tricky.

Many factors interact to influence student achievement, and it is difficult to isolate the contributions of teachers from those of other factors (Assessing Accomplished Teaching: Advanced-Level Certification Programs, 2008, p. 25). The complex questions that education researchers and policymakers are grappling with include:

How can the effect of a teacher on student learning be isolated, when so many external influences promote or hinder student learning?

::::::: How can the limitations of existing student assessment data be accounted for while taking advantage of the wealth of data they provide?

::::::: What other kinds of evidence—for example, observations of teacher practice, samples of student work, and evidence about students (including attendance data and information about their specific learning needs) — can be used to inform valid and reliable measures of teacher effectiveness?

::::::: How long should a teacher be expected to work with a group of students before it is reasonable to expect evidence of learning gains?

As the task force examines the best evidence, theories, and ideas to orient the evaluation of teachers around their contributions to student learning, we are aware that these measures almost certainly will never be perfect. But incorporating them into the National Board Certification process in appropriate ways would be a remarkable advance for education, as well as for credentialing in general.

The remaining sections of this paper describe how student learning is reflected in the NBPTS certification process; assess traditional and alternative measures of student learning for teacher evaluation; and provide a set of recommendations for how NBPTS and other major stakeholders, including the federal government, states, and the philanthropic community, can use measures of student learning to assess teacher effectiveness in increasingly valid ways. ★

Evaluating How Student Learning Is Reflected in the Current NBPTS Certification Process

As a starting point, it is important to assess how the National Board for Professional Teaching Standards includes measures of student learning in the current National Board Certification process, which requires teachers to demonstrate multiple examples of student progress and evidence of whole-class and small-group discourse, along with teacher practice.

National Board Certification is a voluntary three-year certification process. Teachers report an investment of up to 400 hours in reading and understanding the National Board's core propositions and standards, completing four portfolio entries, and sitting for a three-hour assessment administered at a secure testing center. Approximately 40 percent of candidates certify in their first year; 60 percent certify by the end of the second year; and approximately 70 percent certify by the end of the three-year process.

The NBPTS program is a three-tiered process, including a set of core propositions for all teachers; a common set of accomplished teaching standards specific to each content field; and a set of cutting-edge, evidence-based assessments specific to the field that certify what accomplished teachers know and do. Integral to certifying a teacher as accomplished is providing evidence of that teacher's impact on student learning.

Core Propositions

The National Board's framework for accomplished teaching is set forward in its 1989 publication, *What Teachers Should Know and Be Able to Do*. The five core propositions articulated in this publication serve as the foundation for all of the National Board's standards and assessments (see chart, following page). The core propositions define the level of knowledge, skills, abilities, and commitments that accomplished teachers must demonstrate.

The Five Core Propositions

1   **Teachers are committed to students and their learning**.

> Teachers recognize individual differences in their students and adjust their practice accordingly.
>
> Teachers have an understanding of how students develop and learn.
>
> Teachers treat students equitably.
>
> Teachers' mission extends beyond developing the cognitive capacity of their students.

2   **Teachers know the subjects they teach and how to teach those subjects to students**.

> Teachers appreciate how knowledge in their subjects is created, organized, and linked to other disciplines.
>
> Teachers command specialized knowledge of how to convey a subject to students.
>
> Teachers generate multiple pathways to knowledge.

3   **Teachers are responsible for managing and monitoring student learning**.

> Teachers call on multiple methods to meet their goals.
>
> Teachers orchestrate learning in group settings.
>
> Teachers place a premium on student engagement.
>
> Teachers regularly assess student progress.
>
> Teachers are mindful of their principal objectives.

4   **Teachers think systematically about their practice and learn from experience**.

> Teachers are continually making difficult choices that test their judgment.
>
> Teachers seek the advice of others and draw on education research and scholarship to improve their practice.

5   **Teachers are members of learning communities**.

> Teachers contribute to school effectiveness by collaborating with other professionals.
>
> Teachers work collaboratively with parents.
>
> Teachers take advantage of community resources.

Common Standards for Accomplished Teaching

The NBPTS program develops common standards for accomplished teaching that teachers must demonstrate to become certified. Grounded in the five core propositions, field-specific standards articulate the actions that accomplished teachers take to advance student learning. NBPTS has developed content standards for 25 certification areas that represent 16 content fields and six student developmental levels.

Assessments of Accomplished Teaching

Aligned with the core propositions and standards, evidence-based assessments require teachers to demonstrate their practice by providing evidence of what they know and do, while honoring the complexities and demands of teaching. These assessments validate the practice of individual teachers seeking National Board Certification, and, in turn, are validated by research that has identified specific propositions and teaching practices that contribute to student learning. Teachers respond to six assessment exercises designed to tap their content knowledge in ways that distinguish accomplished practice. They also develop four portfolio entries that represent an analysis of their classroom work as it relates to student learning and teacher practice.

Mastery of the content knowledge that contributes to accomplished teaching in a teacher's field— *what the teacher knows*—is assessed by means of a computer-based assessment consisting of six individual, 30-minute exercises administered at a secure testing center. This knowledge base exceeds the upper limits of licensure evaluation instruments.

The four classroom-based portfolio entries require teachers to demonstrate their teaching practice—what they do—and are closely integrated with student learning. In each of the 25 certificate areas, teachers must provide three classroom entries with written commentary and reflection:

⁝ A classroom-based entry with accompanying examples of student work over time, from a minimum of two students with different learning profiles;

- A classroom-based entry that demonstrates whole-class discourse and learning;

- A classroom-based entry that demonstrates small-group discourse and learning; and

- A documented accomplishment entry that provides evidence of the teacher's accomplishments outside the classroom and how that work impacts student learning.

The videotaped and written elements of the portfolio are designed to evoke evidence that demonstrates teachers' (1) effective practice resulting in student learning, (2) mastery of the five core propositions, and (3) mastery of the standards in their content field. The videos require teachers to demonstrate an accomplished level of critical thinking and performance, reflecting the complex and multidimensional nature of teaching and learning. These classroom demonstrations also provide evidence of the effectiveness of the teachers' interactions with students and the students' involvement and learning. The written commentary allows the teacher to describe, analyze, and reflect on his or her instruction and the students' learning.

By combining evidence of student learning and examples of teaching practice with the teachers' analysis of that practice and how it connects to student learning, the portfolio process provides a basis for evaluating not only how teachers performed in the limited snapshot of teaching captured in the portfolio entry, but also the extent of their overall mastery of teaching practice.

Through this process, teachers also demonstrate how they transform the core propositions into practice. The illustration on the following page shows one strand representing teaching practice as grounded in the five core propositions, while the other represents teachers' impact on the students and their learning. When a teacher is accomplished, the double helix is tightly structured.

In order to gauge teaching effectiveness, National Board scorers—all experienced subject-level teachers—examine teachers' classroom interactions with students (provided in the video) and their understanding of how specific lessons serve the goals of student learning (provided in the written materials).

The Architecture of Accomplished Teaching:

WHAT IS UNDERNEATH THE SURFACE?

6TH
Set new high and worthwhile goals that are appropriate for these students at this time.

5TH
Reflect on student learning, the effectiveness of the instructional design, particular concerns, and issues.

4TH
Evaluate student learning in light of the goals and the instruction.

3RD
Implement instruction design to attain these goals.

2ND
Set high, worthwhile goals appropriate for these students at this time, in this setting.

1ST

Your students :
Who are they?
Where are they now?
What do they need and in what order do they need it?
Where should I begin?

Five Core Propositions

Teachers are committed to students and their learning.

Teachers know the subjects they teach and how to teach those subjects to students.

Teachers are responsible for managing and monitoring student learning.

Teachers think systematically about their practice and learn from experience.

Teachers are members of learning communities.

No matter how well classroom resources, subject content materials, or a particular instructional approach are explained or described, effective teaching is demonstrated by student understanding of and engagement with the subject area of instruction. While the written commentary describes how the candidate is effective as a teacher, student work and participation demonstrate the results of effective teaching.

Teachers' attention to student learning is weighed heavily in assessing their level of accomplishment. In assessing the classroom-based portfolio entries, scorers consider the appropriateness of instructional planning, specific classroom instruction, and student assignments. They look carefully at teachers' contextual information and their reflection, noting whether they have appropriate student-learning goals and the ability to make adjustments in order to reach those goals. Teachers who are rated highest demonstrate that they are attentive to student learning and are aware of how their instruction fosters it. When learning or growth has not taken place, the teacher's reflections are of utmost importance. The accomplished teacher is also an accomplished learner, using mistakes to strengthen future teaching practices. ★

Defining Key Concepts

The particulars of how we define learning and teaching have profound implications for our understanding of how measures of student learning and student achievement could be incorporated into the assessment of accomplished teaching. Thus, the task force developed working definitions of student achievement and student learning and specified how the terms relate to accomplished teaching.

Precise definitions of these terms have proven elusive, as each of these concepts has several layers of meaning and nuance. Theoretical conceptions of learning range from the accumulation of bits of information and isolated skills to more holistic notions of critical thinking, reasoning, and communicating within particular disciplines. And as theories of learning vary, so do conceptions of teaching as it relates to student learning. At one end of the spectrum, teaching can be viewed as a linear transfer of knowledge from teacher to student. At the other end, teaching can be seen as mediating, interactive, and interdependent. This latter view of teaching conveys an image of professional, accomplished practice that involves engaging student thinking, continually monitoring and assessing student progress, and adapting instruction to meet student needs.

What counts as learning is by no means universally understood as either a theoretical or technical matter—or even as a matter of priorities and values.

Differences in opinion also exist over the appropriate content and cognitive demands of learning, both within and across academic subject areas. The question of what kinds of skills count as learning is at the core of a range of curriculum debates—do we count memorization of multiplication tables as mathematics learning? Or does learning involve a more constructivist task that requires students to describe how multiplication relates to addition. Or does learning involve some combination of the two approaches?

What counts as learning is by no means universally understood as either a theoretical or technical matter—or even as a matter of priorities and values. While our discussion is largely independent of these debates about priorities and values, we can still offer some basic ideas about how we view these terms.

Student learning and student achievement are closely related concepts. But while the two terms are often used interchangeably, they convey profoundly different ideas, particularly as they relate to teaching. In brief, *student achievement* is the status of subject-matter knowledge, understandings, and skills at one point in time. The most commonly used measure of student achievement is a standardized test. Such standardized assessments measure specific areas of achievement—for example, the extent to which a 3rd grader has mastered the English/language arts standards in his or her state or district—and are best understood as one measure of a subset of a body of skills or knowledge.

Defining Key Terms

Student achievement is the status of subject-matter knowledge, understandings, and skills at one point in time.
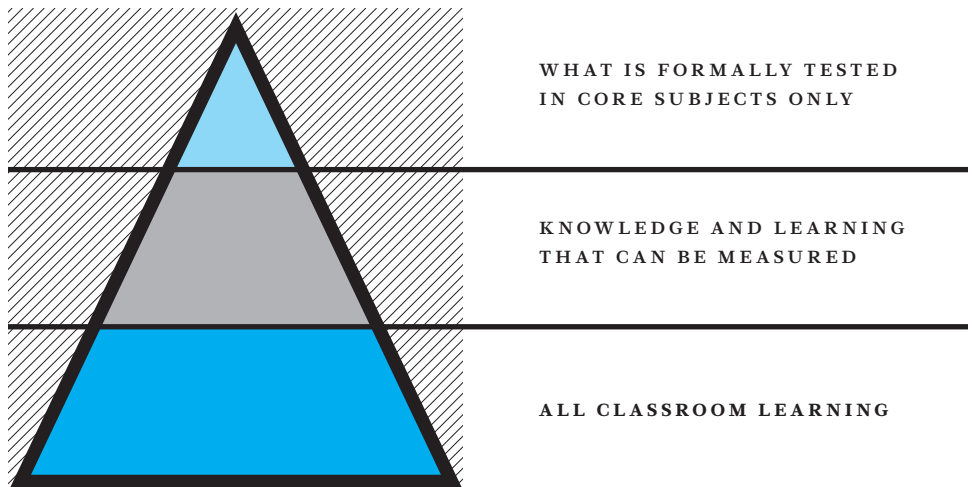
Student learning is growth in subject-matter knowledge, understandings, and skill over time. In essence, a change in achievement constitutes learning. It is student learning—not student achievement—that is most relevant to defining and assessing accomplished teaching.

Accomplished teaching reflects skilled practice and contributes to student learning.

The illustration to the right suggests this relationship. The box represents the broad domain of skills, learning, and knowledge we expect students to know and be able to do. The shaded triangle reflects the considerable—but still incomplete—portion of what students are expected to know that can actually be measured by different means. The bottom of the triangle ⬤ shows the wide base of learning that occurs in any given classroom, while the middle section ⬤ reflects the narrower—but still substantive—range of knowledge that potentially can be measured through a range of assessments and activities by a teacher in the classroom. The top of the triangle ⬤ represents the extent of what is actually measured by formal, wide-scale testing, which typically only covers core subjects such as language arts, math, and, in some cases, social studies and science.

From Learning to Measuring

WHAT IS FORMALLY TESTED IN CORE SUBJECTS ONLY

KNOWLEDGE AND LEARNING THAT CAN BE MEASURED

ALL CLASSROOM LEARNING

In other words, what is tested does count, but much of what counts cannot be tested. Achievement will always be larger than a single test and is not specific to any particular assessment. Teachers must monitor achievement regularly using a variety of formal and informal assessments for both individual students and the class as a whole.

Student learning is the growth in subject-matter knowledge, understanding, and skills over time. In essence, it is an increase in achievement that constitutes learning. Central to this notion of learning as growth is change over time. Knowing whether student learning has occurred, then, requires tracking the growth in what students know and can do. It is only by comparing student mastery at successive points in time that the nature and extent of learning can be gauged. Student learning is also reflected in a broad array of outcome measures, including attendance, participation, engagement, and motivation.

Measures of learning also vary. One major source of this variation is the different ways in which state standards introduce concepts of varying difficulty at different times. Indeed, a recent National Center for Education Statistics (NCES) study (Mapping State Proficiency Standards Onto

NAEP Scales: 2005–2007) compared performance standards across states and found tremendous variability. States have very different means of identifying when students have made certain gains, including meeting state standards or definitions of proficiency. Using the NAEP (National Assessment of Educational Progress) as a common measure, the NCES study showed that students who made the same progress over time may or may not meet required performance standards, depending on the state in which they live.

How do these concepts relate to teaching? Because student achievement reveals what pupils know, understand, and can do at one point in time, it can be useful for identifying gaps between what students are expected to know and what they actually do know. Teachers can use student achievement information to focus instruction on areas where students are struggling.

············
············

What is tested does count, but much of what counts cannot be tested.

············
············

But there are limits on what achievement information can do to shape instruction. By itself, student achievement reveals little about how to address those gaps. And achievement data alone are not useful in assessing teacher performance, as it is impossible to attribute the influence of the teacher to a single snapshot of student achievement. Student achievement reveals nothing about how that achievement has changed in the short or long term, or what factors—related to instruction or other influences—contributed to that achievement.

In short, it is student learning—not student achievement—that is relevant to defining and assessing accomplished teaching. Drawing conclusions about teacher performance requires an analysis of the influence of teacher instruction on how a student progresses. Analyzing the impact of teacher instruction on students requires a careful, sequential examination of student achievement prior to instruction, the nature and quality of instruction developed and delivered to help students learn, and student achievement after instruction—that is, examining student learning over time as it relates to the work of a teacher.

The causal inference that gains in student achievement are due to a teacher is not easily justified. As is true of many causal inferences, there are many competing explanations for student learning beyond teacher effectiveness. Differences in learning for students with different teachers could be due to differences in parental support or peer groups or a variety of other factors, including tutoring, attendance, individual students' ability to teach themselves or read independently, and the contributions of previous teachers and those working on similar skills in other subjects. Although it is not possible to rule out all possible alternate explanations, the notion that teachers are a critical causal element can be supported by analyses that control for other factors such as parental socioeconomic status and student attendance—and by direct information about teacher performance in the classroom.

We have established how we view learning in the context of teaching, but we still need to define specifically what we mean by accomplished teaching. Taking our cue from an influential paper that considers the conceptual subtleties of defining quality teaching, we assert that accomplished teaching reflects skilled practice and produces student learning (Fenstermacher and Richardson, 2005). First, accomplished teaching meets high professional standards for instructional method and content—that is, it reflects skilled practice and places a value on how something is taught. It is important to note that value is also placed on whether something has been achieved through the act of teaching—that is, whether students learn. Accomplished teaching involves teaching practice that is grounded in an understanding of how to facilitate student learning and that leads to growth in student understanding over time.

Though this definition embodies a tight coupling of teaching and learning, it is important to underscore the point that teaching is not the only determinant of learning. The environment for learning, the engagement of the learner, and the existing resources and opportunities to learn are all influential in shaping student learning. These outside influences on learning also shape how teachers respond to student needs.

These ideas form the conceptual basis for how the task force views a teacher's "value-added" to student learning. In order to gauge what a teacher contributes to progress observed in students over time, we must both look to direct measures of student learning and relate the teacher's practice to student learning. ★

Essential Criteria for Using Large-Scale Standardized Assessments in Teacher Evaluation Systems

We support the use of large-scale standardized assessment results as one measure in the certification process if they enable the calculation of a meaningful gain in student learning. Many state tests currently do not meet the criteria, even though the obstacles to do so are not insurmountable. Here we sketch the minimum conditions that would need to be present in order to make these inclusions feasible and, therefore, acceptable.

Curriculum-related scale with equivalent unit of measure along a considerable continuum of achievement. To claim that a teacher influenced student learning, assessment measures must be closely aligned with standards and must measure student performance at the level where a student actually achieves. Vertical scaling is desired, although not necessarily required, to accurately measure gains in student learning.[1]

Information on validity of tests for assessing special populations. A National Board Certification candidate may be teaching a large proportion of English-language learners but may teach in a state whose assessment is not validated for this population; information on validation for different groups of students needs to be available to find such mismatches.

Data system that tracks students and links to teachers. Assessments of a teacher's ability to procure learning in his or her students require longitudinal data. As we have said, learning is about the growth in student understanding over time, and if we are to attempt to attribute that learning to a teacher's instruction, we must have data at multiple points in time as the teacher engages with those students.

....
....

1   Although vertical scaling is desirable for value-added modeling, it has its drawbacks. For example, it does not measure grade level or content standards as well, because testmakers cannot include as much in these measures. On the other hand, tests of grade-level content standards often fail to measure growth for those who are achieving below or above grade level. So there are trade-offs that require consideration.

Alignment. Several states use both state-developed, criterion-referenced tests to monitor student achievement and commercially available, norm-referenced tests to compare the performance of their students with that of other states' students and the nation as a whole. Only a handful of states use only commercially available tests. Typically such tests are augmented or otherwise altered to align them better with these states' curricula. For states that use only commercially available tests, it is advisable to have adequate documentation that the tests are aligned with the state's curriculum. The commercially available tests in use by the various states include the Iowa Test of Basic Skills (ITBS), the Stanford Series, The Otis Lennon School Ability Test, Comprehensive Test of Basic Skills (CTBS), and TerraNova. See *Appendix A* for a list of tests used in each state.

Since the passage of the No Child Left Behind Act, all 50 states and territories (including states such as Vermont that have concentrated on portfolio assessment) have developed assessments that include some multiple-choice questions. Although the law requires the reporting of Adequate Yearly Progress (AYP), which in turn implies annual testing, it is not clear that all states currently test all eligible students annually. In addition to multiple-choice tests, many states' assessments include short-answer and extended-response exercises, including responses to writing prompts, which allows them to assess a wider range of standards and curriculum expectations. Assessments should satisfy some minimal standard of reliability.

Even with the use of standardized tests that meet these criteria, however, teacher evaluation systems will need to incorporate additional evidence of teacher practice in order to correlate any student learning gains with specific classroom activities. This need is all the more critical because gains in student learning are not just the function of the classroom teacher but of many other factors as well, including teaching conditions and supports, past learning experiences, tutors, parents, student attendance and participation, and other external student and family factors. Having better tests will solve some—but not all—of the dilemmas associated with drawing inferences about the effects of individual teachers on student learning. As stated previously, the task force views these challenges as substantial, but not insurmountable, particularly if policymakers carefully evaluate the strengths and weaknesses of varying approaches to assessing student and teacher performance.

Assessing Traditional and Alternative Measures of Student Learning for Teacher Evaluation

To reiterate the task force's core argument, the question is not whether student learning should be used to evaluate teacher performance, but how. In this section, we tackle the policy debate about whether and how the traditional measure of student learning—standardized achievement tests—should be used as an indicator of effective teaching. We then set out essential criteria for including the results of large-scale standardized student assessments in the evaluation of teacher practice and conclude with principles to guide the ongoing work to improve such measures.

Traditional Measures: Standardized Achievement Tests

On the surface, it seems reasonable to gauge teachers' effectiveness by what their students know and can do. And because of the annual testing requirements included in the U.S. Elementary and Secondary Education Act (ESEA), commonly known as No Child Left Behind, it seems logical to use those results to assess teacher performance based on how well their students perform on these large-scale standardized tests of academic achievement.

This is a compelling idea and one that holds considerable sway in the current policy discourse. To comply with the accountability focus, student test data are generally easily accessible to researchers and policymakers. Because many of these tests are administered state- or district-wide, they generate a wealth of data across classrooms, schools, and even districts. As a result, these measures can be used to make a number of informative comparisons that are helpful in assessing the relative effectiveness of teachers in producing gains on the tests. And in the case of the National Board, there is a strong intuitive case to be made that the students of National Board Certified Teachers ought to be doing well on these tests.

Looking beneath the surface raises questions, however. A wide range of standardized measures exist, and assessments other than those administered under the auspices of high-stakes, state-level testing can offer more valid information for particular purposes. And all standardized measures, including the emerging "value-added" measures that have become the dominant growth

models in use, must continually be evaluated against each other and against other alternative means of assessment through the prism of a wide range of issues, including alignment, metrics, inclusion and accommodation, sensitivity, breadth, and scaling and equivalence, which we examine in more detail below.

Alignment. It is a commonly accepted principle of standards-based reform that student assessments, curriculum, and instruction are all aligned with each state's respective student learning goals. While states that have developed (or commissioned) their own assessments to monitor student achievement have satisfied this requirement, states that use nationally normed tests developed by other testing organizations may or may not have taken appropriate steps to ensure that the test is aligned with their stated curricular and instructional objectives. While all states have evaluated how well their tests are aligned with academic-content standards and submitted the results of these studies to the federal government, the degree of alignment varies substantially from state to state.

Since these tests are designed to assess proficiency in core subjects in light of specific standards, they reflect the nature and level of the knowledge and skills called for in those standards. Although the nature and rigor of standards varies, state standards—and the tests designed to assess proficiency against them—generally target basic skills. If teacher evaluation systems are intended to identify educators who demonstrate skill in advancing higher-order thinking and problem-solving skills among their students, it would not make sense to use measures of basic skills as indicators of accomplished teaching.

Metrics. As we suggested in defining accomplished teaching, the only adequate way to capture the effect of a teacher on student learning is to use a measure of student learning—not achievement—so changes in student understanding can be attributed to instruction. Relying solely on year-end test results to evaluate teachers is invalid and would be especially unfair for teachers who teach students who enter their classrooms seriously behind their cohorts. Again, there must be a measure of entering achievement as well as a measure of end-of-year achievement—that is, a measure of student learning.

One implication of this requirement is that incorporating student learning into the assessment of teachers who teach in the very early grades will be problematic. Although some states (including Arkansas, Idaho, and South Carolina) administer large-scale, state-wide tests of reading readiness, psychomotor development, and other school-relevant developmental skills to children in K–2, most do not. Since there is often no large-scale measure of entering achievement in 3rd grade, there may not be a measure of student learning to use for those teachers. Another issue is identifying the appropriate teacher: Is it the teacher of record, the reading teacher, math specialist, or a collaborative team? The same issue holds true at the other end of the K–12 spectrum, as almost no states have value-added measures of student learning in all high school subject areas.

Value-added Models. In recent years, there has been substantial interest in the use of value-added models to analyze student test scores in order to estimate the contributions that specific teachers and schools make to the growth in student learning.

While the term "value-added" is used mainly in conjunction with K–12 student achievement, it could be applied more broadly to other student outcomes, such as graduation rates. Value-added models differ in their degree of sophistication, but all are based on the same core premise. The models use prior information (for example, test scores and/or other student data) to estimate expected outcomes for each student at the end of each year. Those expected outcomes are then compared to actual student outcomes. The difference between the actual and expected outcomes is the "value-added" by the teacher, school, or program that is the focus of the analysis.

To see the potential importance of the value-added concept, consider that traditional standardized measures assess schools based on the percentage of students who are proficient. Implicitly, they assume that students in every school are the same at the beginning of the school year, even though we know that students come to school—even to kindergarten—with varying levels of readiness. This is why value-added measures provide better information about what schools contribute to student learning than do snapshots of student achievement that fail to account for these external influences on student achievement.

The reason the idea of value-added is applied mainly to K–12 student achievement is that it is easier to estimate each student's expected outcomes when we have measures over time for each individual student. Student scores are highly correlated over time, so if we wanted to predict a student's 6th-grade math achievement, his or her 5th-grade math achievement would be a good predictor of 6th-grade scores. Models that take into account more information about student achievement (for example, achievement in both reading and math in 3rd, 4th, and 5th grades) are more defensible than are models that use only a single prior achievement score because the additional prior information improves the accuracy of the estimated outcome.

While evaluating teachers, schools, and programs with value-added models is almost certainly better than looking at snapshots of student achievement at a single point in time, it is not without challenges. The accuracy of value-added measures is lessened by achievement tests that do not yield equal interval scales and are unable to account for school-level factors or unmeasured student characteristics influencing each teacher's success. Research on the "teacher effect" in value-added models suggests that measures of individual teachers' performance are sensitive to the specific statistical methods and ways in which student achievement is actually measured, including the alignment of the assessment to the curriculum and the students being assessed.

And none of the models can provide conclusive evidence that any effects are attributable only to differences in teacher effectiveness. Value-added measures of changes in student achievement are a function of many things in addition to the contribution of any individual teacher, including other teachers who work with the student; school-level resources and variables, including class sizes, libraries, computers, texts, and the presence of facilitators and other support personnel; the contributions of teachers the student has had in the past; curriculum decisions; and personal variables impacting each individual student's ability to learn, including home and health factors. The practical significance of these external factors for value-added measures is largely unknown, and some are indirectly accounted for because they are related to the prior test scores that form the basis of value-added. But there is little doubt that value-added measures do not account for all of these factors, other than teaching, that influence student learning.

The strengths of value-added models also vary based on their specific purpose. Estimating the impact of individual teachers on student achievement is particularly imprecise because each teacher has a relatively small number of students. Estimating school value-added is somewhat easier for the same reason; there are more students in each school than in each teacher's classroom, yielding more information on which to base the value-added estimates. Least controversial of all is using value-added to assess large-scale education programs. NBPTS is an example of a program that has been evaluated using value-added methods to assess the effectiveness of National Board Certified Teachers. These results are more trustworthy because the studies are based on patterns observed in thousands of teachers and schools, allowing researchers to draw conclusions with a high degree of confidence.

In the case of teacher value-added, there is a significant difference between using such measures to inform professional development and using them in evaluation and compensation systems. The higher the stakes attached to any measure, the higher the standards we must expect those measures to live up to.

Inclusion and Accommodation. The number of students with limited English proficiency (LEP) varies widely across states and districts, as do state policies regarding how to test such students (that is, whether and in what way such students benefit from accommodations when taking the test) and how their results are reported for accountability purposes. Using state test data to assess teachers—particularly those who teach large numbers of LEP students—requires careful consideration of how to take into account whether tests are administered and reported for LEP students and whether the tests are valid for those particular students.

Sensitivity. The ability of tests to measure the range of the scale commensurate with a student's ability varies. For example, if the standardized assessment does not have sufficient measurement reliability and validity at a student's actual achievement level, gains in that student's learning may be difficult to detect. While most tests have better measurement validity in the middle, as opposed to the highest and lowest levels of student achievement, tests that are too hard or too easy are problematic for measuring achievement in certain groups of students. Too much measurement error in certain ranges could also lead to an inability to detect student growth for stu-

dents whose ability does not match the ranges for which the test is most sensitive. States that use tests focused only on grade-level standards, for example, may not be measuring student learning adequately for those who are achieving well below or well above grade level.

Breadth. Given the prevailing requirements of measuring Adequate Yearly Progress (AYP), student test results are typically available only for grades 3 through 8 and in reading/language arts and mathematics (and, increasingly, science). Furthermore, these tests are often not scaled in a way that permits the measurement of growth from year to year. Estimates suggest that, even in states with vertically scaled tests, only about 30 percent of K–12 teachers would have such student test scores available to develop measures of student growth in achievement for teacher evaluations. For National Board Certification in particular, questions of fairness and comparability would need to be addressed if such measures were used in only a handful of NBPTS areas.

Scaling and Equivalence. State and district achievement tests differ not only in the scaled scores used to report results to the public but also in the content of what is tested. As a result, measures of student learning derived from such tests are not comparable across states or districts. Converting growth measures to "effect sizes" can address the scale problem, but doing so does not account for differences in content. Although considerable content overlap across states is to be expected in any given level-subject combination (for example, 4th-grade mathematics), complete content overlap—that is, equivalence—is neither attainable nor, some would argue, desirable.

Also, reaching full equivalence of achievement scales across districts and states may not be strictly necessary. From the very beginning, the National Board has applied the same evaluation rubric to teachers who teach in non-equivalent circumstances and whose students show different levels of growth. To require such equivalence for the inclusion of standardized achievement test data in the certification program is, therefore, not entirely consistent with past NBPTS practice. As long as all teachers are evaluated based on what they are expected to teach in their respective states, this does not appear to be a significant concern.

Integrating Indices of Student Learning with Teacher Assessments. There is also the matter of precisely how student learning, as reflected in gains on standardized tests, would be figured into National Board candidates' overall scores or used in other teacher evaluation systems. The possibilities are too numerous to list. At one end is a simple dichotomy: Candidates are awarded a specified score increment if their students' mean or median growth—when appropriately scaled—exceeds some pre-specified amount. One of many possible variations on this theme might be a specified candidate score increment if a given percentage of students exceeds a certain scaled gain. At the other extreme are multi-point or graduated systems in which candidates obtain higher scores depending upon the mean or median growth of their students.

The way in which student learning would be incorporated into these scores is not a simple matter of psychometric taste. By encompassing a full range of standards and a critical analysis of teaching practice, the National Board Certification process focuses candidates' attention on a broad spectrum of student learning. Narrower measures of student learning that are part of teacher assessments, by contrast, would directly affect what teachers do in the classroom. For example, if mean student growth is the critical statistic, teachers might concentrate their attention on the students they expect will show the largest gains. If a specified score increment is awarded if a given percentage of students exceeds a certain scaled gain, teachers will concentrate their instructional efforts on students perceived to be closest to the critical score. In short, how student learning is explicitly incorporated into evaluation and certification decisions could affect teacher behavior and decisions in the classroom in ways that may not always be instructionally sound.

Linking Student Records to Teacher Assessment. As previously noted, one result of E S E A is that every state now has some form of annual testing in grades 3 through 8 that includes standardized, multiple-choice testing. Linking that student achievement data to individual teachers, however, has proven difficult. The problems encountered are legion: multiple student I D s attached to the same student, multiple students with the same I D s, students who pop in and out of the database in seemingly random fashion, and on and on. Rarely do student matches over even a single year exceed 80 percent.

Beyond the technical issues, many schools place multiple teachers in the same classroom, making the link to a single teacher ambiguous. For example, some teachers have the benefit of adult aides, while others do not. Some schools also have mature and vigorous "pull-out" programs providing one-on-one tutoring to students with disabilities or other instructional needs. Finally, many districts and schools within districts are plagued by rampant student transiency. Indeed, the student-teacher link may well be the most problematic hurdle in including standardized testing in the mix of student learning. Tight coordination with state offices of assessment will be required, but if history is any guide, there is no guarantee that such coordination will result in credible and comparably complete links.

We view all of these challenges as substantial but not insurmountable. And practical difficulties or conceptual challenges in no way should be taken as excuses for inaction or justifications for troubling current practices. We raise these concerns, not as reasons to resist current efforts to systematize teacher evaluation, but as important tasks in ensuring that we do evaluate as effectively as possible. On pages 32 and 33 we set forth some essential criteria that state tests should meet in order to be used as valid and reliable measures of teacher effectiveness.

Whether for teachers or students, all scoring systems contain difficult problems that must be carefully considered and analyzed. Given the high-stakes character of National Board Certification, it is worth emphasizing that any system, no matter how well thought out, will be imperfect and will need to be constantly monitored and weighed against other alternatives. But these cautions are meant as just that: cautions. As in the case of the larger policy context, they are not meant to thwart continued innovation and improvement in the National Board Certification process.

Alternative Measures Currently Used in Teacher Evaluation and the Assessment of Teaching Practice

The task force explored several examples of approaches that ground teacher evaluation, credentialing, and incentive structures in student learning, including Oregon's Teacher Work Sample (TWS), The Renaissance Teacher Work Sample, Denver's Professional Compensation

for Teachers Program ("Pro-Comp"), and Arizona's Career Ladder. While the task force does not endorse any specific approach and believes much work needs to be done in this area, these examples incorporate such elements as including evidence from the classroom and measures of student learning as part of a broader series of instruments used to evaluate teacher effectiveness.

Along with these integrated approaches developed by districts and states, the task force surveyed a series of emerging instruments that more specifically assess teaching practice, including the Classroom Assessment Scoring System (CLASS), the Learning Mathematics for Teaching (LMT) Project, and the Protocol for Language Arts Teaching Observations (PLATO). (See *Appendix C* for details.) While these instruments serve as a source of ideas about expanded or alternate methods of incorporating direct measures of classroom practice into broader evaluations of teacher effectiveness, they do not address incorporating measures of student learning into assessments of teaching practice.

Remaining Challenges

Along with the work of the National Board, these emerging approaches and instruments represent some of the most forward-thinking work in the field to ground teacher evaluation, credentialing, and reward structures in discrete examples of student learning and teacher practice. At the same time, it is clear to us that there is much work to be done in this field.

To help improve the validity of a range of measures, we draw attention to the possibility of using student growth measures in research as a validation of the kinds of practices that ought to make up performance assessment tasks and other measures. Specifically, as the small but growing body of research becomes more prevalent, it will enable identification of "instructional correlates" that predict value-added to student learning. The practices that have been shown to predict student learning could be included and heavily weighted in performance assessments, while teaching practices that lack this predictive validity would be weighted less or dropped. ★

Recommendations

Recommendations for Student Assessments

To build on the promising elements indentified in the previous sections of this white paper, the task force has drawn out a series of principles for selecting or developing student assessments that are used to evaluate teacher practice should:

1   **Be aligned with the curriculum and student learning goals a specific teacher is expected to teach.** Measures of student learning must reflect the specific content of what is expected to be taught and must be explicitly aligned with the curriculum elements for which individual teachers are responsible. This principle also recognizes the importance of identifying the specific teacher or teachers responsible for gains in student learning, particularly because learning is a cumulative process, with previous teachers and learning experiences playing significant roles.

2   **Be constructed to evaluate student learning**—that is, performance at two or more points in time, rather than a snapshot of student achievement, so that changes in students' understanding and performance can be substantially attributed to instruction. This principle applies with equal force to standardized quantitative measures and more qualitative measures of student learning, such as portfolios of student work—both of which must focus on the gains in learning students have realized over the period during which a teacher provided instruction.

3   **Be sensitive to the diversity of students,** including those with special needs or limited English proficiency, as well as gifted or high-achieving students. Assessments used to evaluate teachers must be valid for the student populations they teach.

4   **Capture learning validly and reliably at the students' actual achievement level.** Measures should be evaluated continuously to determine the extent to which they address the principles of alignment with the range of knowledge and skills to be measured and the ability to capture student learning across the diverse learning needs and backgrounds outlined in this paper.

5 **Provide evidence about student performance and teacher practice that reflects the full breadth of subject-matter knowledge and skills that are valued**. This recommendation addresses the need to identify the extent to which a teacher's practices are connected to and influence student learning. Linking these measures enables a rich and nuanced assessment of on-the-ground practice in context and can capture the complexities of the effects of teaching on student learning over time.

Recommendations for Teacher Assessment Systems

The same principles that guide assessments of student learning should apply to evaluations of teacher practice. As a response to the evolving conditions in assessment and policy, we have translated these broadly accepted principles to specific recommendations to guide practice. The task force recommends that assessments or evaluations of teaching practice:

1 **Be grounded in student learning, not student achievement**. This recommendation applies with equal force to standardized quantitative measures as well as more qualitative measures. A single achievement measure, by contrast, reveals only a snapshot of student understanding at one point in time—and very little about the teacher's influence. The only defensible way to determine teacher effectiveness is to focus on the gains that students have realized over the period during which the teacher provided instruction. For example, an analysis of student work before and after a teacher's instructional intervention provides the conceptual basis for inferring that the teacher had a positive influence on individual student learning.

2 **Employ measures of student learning explicitly aligned with the elements of curriculum for which the teachers are responsible**. This recommendation emphasizes the importance of ensuring that teachers are evaluated for what they are teaching. For example, the selection of the assessment must reflect the specific content being taught, including higher-order thinking and concepts. Tests may need to be differentiated to address the needs of the groups of students being taught, including students with disabilities or language-acquisition needs.

3 **Strive to attribute student growth to the teachers responsible**. This recommendation underscores the importance of unambiguously attributing gains in student learning to a teacher's contribution to students' learning—and to the specific teacher responsible for the gains. For instance, value-added systems today face considerable challenges in distinguishing between instruction a classroom teacher provides and instruction provided by a resource specialist. In evaluating or recognizing teacher performance, identifying the correct teacher matters. This issue will become increasingly pronounced as districts and schools employ innovative staffing configurations such as team teaching, flexible grouping, and virtual delivery. The process by which teachers associate learning gains over time with their instructional plans and strategies also allows them to adapt their teaching practices to address specific student needs.

4 **Establish the link between student learning and teacher practice**. This recommendation addresses the need to identify the extent to which a teacher's practices are connected to and influence student learning. Well-configured systems ought to consider teacher practice to ensure that it is consistent with measures of student learning. Linking these measures enables a rich and nuanced assessment of on-the-ground practice in context and can capture the complexities of the effects of teaching on student learning over time. We define accomplished teaching as being a function of both teaching practice and student learning. Evaluation of teacher effectiveness, then, needs to include measures of both. The teacher work sample initiatives highlighted in *Appendix D* offer one illustration of how multiple measures can be considered in enabling in-depth assessments of a range of competencies of accomplished teachers—for example, the quality of the teachers' assignments and the way they assess, plan, adapt, and provide feedback in relation to individual student work over the course of a lesson or unit. These measures can also be flexible, in that a teacher could choose a range of outcomes related to learning (for example, assessment information about student mastery in core subject areas or homework completion) as well as a range of assessment tools, including teacher-developed measures.

5 **Use measures that, to the greatest extent possible, reflect the full curriculum, the full scope of a teacher's responsibilities, and the full domain of skills and competencies students are expected to develop.** Measures should be evaluated continuously to determine the extent to which they address the principles of alignment with the range of knowledge and skills to be measured and the ability to capture student learning across the diverse learning needs and backgrounds outlined in this paper.

Recommendations for NBPTS

While no approach is perfect, these recommendations are intended to serve as guidelines in designing teacher assessment systems that reflect student learning and improve teaching practice. To that end, the task force believes that National Board Certification should ultimately be a measure of how accomplished teachers are contributing to student learning. While the National Board Certification process already requires teachers to demonstrate multiple examples of student learning, we recommend that NBPTS:

1 **Explore strengthening the extent to which student learning is systematically evaluated in each of the 25 certificate areas.** The task force recommends that the National Board be more precise about the nature of student work submitted in the portfolio process so that the work measures student learning more accurately in relation to teaching practice. This recommendation includes urging the National Board to strengthen evidence of student learning in each certification area, including systematic representations of learning and high-quality assessments wherever they are available.

One vision of an authentic student learning portfolio task—which takes its cue from the promising practices outlined in the previous section—would require candidates to think about student learning in everything they do and show that they produced learning over time by assembling a collection of evidence that demonstrates student learning. Teachers should be able to demonstrate mastery of student learning performance tasks, including, but not limited to:

- Assessing and analyzing student work before instruction. Accomplished teachers need to know how to gauge where students are before developing and teaching a lesson or unit. They should be able to clearly articulate the criteria used to select the assessment tool and how that tool was used to evaluate student work. Accomplished teachers then craft lessons or units that build on, and address deficiencies in, student understanding. They develop instructional plans that begin where students are and move toward where they need to be.

- Providing instruction based on student work. Accomplished teachers deliver lessons as planned, although they make adaptations along the way based on an ongoing assessment of student learning during the course of instruction.

- Assessing and analyzing student work after instruction to reflect on instruction. Accomplished teachers gauge where students are after each lesson or unit to determine whether and how learning has occurred, and then evaluate their own success in delivering excellent instruction in light of that evidence. This evaluation should drive subsequent planning that supports the next steps in student learning.

- Providing feedback to students based on their progress to guide student reflection and revision. Accomplished teachers show that they engage students in ways that reflect students' growth in understanding.

Candidates should also be required to continue to provide evidence of the following:

- Growth in student learning over time for a handful of students (at least two, and preferably as many as five) by showing student work samples prior to instruction and again after instruction, demonstrating teacher influence on particulars of individual student mastery and growth.

- Growth in student learning over time for the whole class by showing an aggregate measure of student understanding prior to instruction and demonstrating teacher influence on the growth of the class as a group.

- Teacher assignments requiring students to engage in complex higher-order problem-solving skills, which ensure that teachers are engaging their students in ambitious work and not sacrificing the quality of student assignments in order to obtain a favorable student learning assessment.

2   **Explore adding additional evidence of student learning, created by teachers and derived from broader assessment measures, to the basket of evidence currently used in the National Board Certification process**. Following models such as those explored in this paper, NBPTS could, for example, develop criteria for using standardized assessment results from the school, district, or state level in programs that tie teacher evaluation to student learning. It could also require teachers to submit, on a pilot basis, existing state or district assessment data, where aligned, valid, and available, as well as alternative measures of student learning in school districts and subject areas to augment standardized data or where such standardized data are not available. Where these measures are used, they should be evaluated in conjunction with other data about the characteristics of students, the context of instruction, and the teachers' practices, so that inferences can take into account the factors that would influence score gains and attributions about their sources.

Many technical problems must be resolved before such measures can be used validly and fairly in National Board Certification, including matching student records to candidates, addressing inclusion and accommodation issues, curricular alignment, the appropriateness of the test for measuring gains, and defining how student learning indices will actually contribute to candidate scores. However, NBPTS could advance the field and improve the national discourse around teacher evaluation–related policy proposals by developing a list of essential criteria for using state and district test results in programs that tie teacher evaluation to student learning. This published list eventually could serve as a set of standards that candidates must meet in order to include such measures in their portfolios. We have outlined an initial set of criteria on pages 32 and 33.

3   **Continuously monitor research on the impact of teachers on student learning**. As the body of research continues to emerge, NBPTS should continually study the evidence and test the validity of its own standards and instruments.

4   **Through the National Board's research, promote systematic use of methods for evaluating teachers' effectiveness and impact on student learning**. The National Board should

conduct research and share the results with other stakeholders to help inform the use of information and assessments of both student learning and teacher effectiveness.

The possibilities include expanding the nascent research base on the predictive validity of NBPTS portfolio entries to measures of student learning. Such studies could intentionally vary the set of performance tasks candidates are asked to complete in order to assess the degree to which different portfolio assessments and their features—number, type, relative weight—predict teacher effectiveness scores. Another possibility could be funding exploratory research on different ways the National Board might incorporate the value-added notion into its certification processes. We have suggested the possibility of revising one portfolio task per certificate area to include at least one task tied to student growth; a study could help identify others.

5   **Promote the development of teacher skills in designing classroom assessments and interpreting external assessment results, providing appropriate feedback to students, and using measures of student learning as a central indication of accomplished teaching.** These are important aspects of teacher practice that bear directly on how much teachers contribute to student learning. Teachers need to understand how a system of assessments helps to define the framework for their teaching and contribute to a complete portrait of the student as a learner in the classroom. The more sophisticated teacher-created classroom assessments that would result from the development of such skills could become a strong component of the National Board Certification process. These assessments provide a personal, classroom-level connection between student learning data and an individual teacher's practice.

A range of skills is involved in designing classroom assessments and interpreting external assessment results. State, district, and formative classroom-level assessments (for example, end-of-book/course, chapter, teacher-constructed quizzes, portfolios, and diagnostic assessments) are designed to make unique contributions to a teacher's broader understanding of students' strengths and needs, while informing the central element of accomplished teaching. Accomplished teachers need to be informed consumers of each

test available in a system of assessments. This means they need to know and appreciate key design principles affecting the integrity and utility of such assessments, including industry standards for acceptable levels of measurement reliability and validity and the validity of such assessments for student groups with diverse learning abilities, styles, and developmental status.

It is equally important that teachers know how to move from data to data-driven instruction. Accomplished teachers must be able to manage, interpret, and use data to adapt instruction to meet student needs, and then follow up to assess the impact of their instruction. They must demonstrate their understanding of assessment systems as engines that drive improved student learning in the direction schools, districts, and states have specified in their learning standards, objectives, and achievement levels.

To prepare teachers to effectively use a system of assessments at the state, district, and classroom levels, most pre-service teaching programs will need to be augmented to include multiple supervised opportunities. Pre-service teachers will learn about formative and summative assessments. They should apply and discuss what they are learning in supervised classroom situations so they are prepared to work collaboratively with complex, standards-based assessment systems. Comparable improvement of the current teaching force should take place within ongoing, job-embedded professional development allowing teachers to apply their new knowledge to their current work and to learn from the experiences of their colleagues. NBPTS can exercise its considerable voice and vision to bring about such changes. ★

Conclusion

By now, our unwavering support for using student learning as a cornerstone of teacher evaluation should be clear. It should be equally clear that much work needs to be done to research and refine the best ways of incorporating measures of student learning into teacher evaluation systems.

As new approaches emerge, this report underscores the need for educators and policymakers to combine multiple measures of student learning with a comprehensive approach to measuring accomplished teaching practice and student learning. The task force believes that the National Board can play a critical role in the broader policy conversations on measuring teacher performance by communicating the broad principles that guide its systems and measures, as well as the approaches needed to better gauge teachers' roles in student learning.

For nearly a quarter-century, NBPTS has played a leading role in identifying what both accomplished teachers and high-achieving students are expected to know and be able to do. We applaud the current emphasis on identifying, rewarding, and placing teachers based on their effectiveness in promoting student learning and hope this paper might help both the National Board and the national policy community advance these efforts in credible, thoughtful ways.

The Student Learning, Student Achievement Task Force,
National Board for Professional Teaching Standards

ROBERT LINN

LLOYD BOND

PEGGY CARR

LINDA DARLING-HAMMOND

DOUGLAS HARRIS

FREDERICK HESS

LEE SHULMAN

Appendix A.
Summary Table of State Testing in Elementary and Middle School

State	Test	Grades	CRT	NRT	MCT	ER
A L	State/Stanford10	3–8	✔	✔	✔	✔
A K	State/Terra Nova	3–9/5,7	✔	✔	✔	✔
A Z	State/Stanford10	3–8	✔	✔	✔	
A R	State/Stanford10	3–8	✔	✔	✔	✔
C A	State CST	2–11	✔		✔	✔
C O	State	3–10	✔		✔	✔
C T	State	3–8	✔		✔	
D E	State	2–10	✔		✔	✔
F L	State/Stanford 10	3–10	✔	✔	✔	✔
G A	State	1–8	✔	✔*	✔	✔
H I	State/Terra Nova	3–8, 10	✔	✔	✔	✔
I D	State	3–10	✔		✔	✔
I L	State/Stanford 10	3–8	✔	✔	✔	✔
I N	State	3–8	✔		✔	✔
I A	ITBS	K–8		✔	✔	
K S	State	3–8	✔		✔	✔
K Y	State, ITBS	3–8/3–7	✔	✔	✔	✔
L A	State/iLeap	1/743	✔	✔	✔	✔
M E	NECAP	3–8	✔		✔	✔
M D	State/Stanford 10	3–8	✔	✔	✔	✔
M A	State (MCAS)	3–8, 10	✔		✔	✔
M I	State	3–8	✔		✔	✔
M N	State (MCA II)	3–8, 10	✔		✔	✔
M S	State	3–8	✔		✔	✔
M O	State (MAP)	3–8	✔	✔	✔	✔

CRT	Criterion-referenced test (or Standards-Referenced Test)
NRT	Norm-referenced test
MCT	Multiple-choice Test
ER	Extended response test (including "short answer," writing, etc.)
*	Optional
†	Home Schooling Only

The table is read as follows:

The Montana State test is administered annually to grades 3 through 8; the ITBS is administered annually to grades 4 and 8; and the assessment includes both norm-referenced and standards-referenced tests, as well as multiple-choice and extended-response questions.

State	Test	Grades	CRT	NRT	MCT	ER
MT	State	3–8, 10	✔		✔	✔
NE	State	3–8, 10	✔		✔	✔
NV	State	3–8	✔		✔	✔
NH	NECAP	3–8	✔		✔	✔
NJ	State	3–8	✔		✔	✔
NM	State/Terra Nova	3–8, 11	✔	✔	✔	✔
NY	State	3–8	✔		✔	✔
NC	State	3–8	✔		✔	
ND	CTB/McGraw Hill	3–8	✔		✔	✔
OH	State	3–8	✔		✔	✔
OK	State	3–8	✔		✔	✔
OR	State	3–8	✔		✔	✔
PA	State	3–8	✔		✔	✔
RI	NECAP	3–8	✔		✔	✔
SC	State/Terra Nova	3–8	✔	✔	✔	✔
SD	State/Stanford10[†]	3–8, 11/2 4, 8, 11	✔	✔	✔	✔
TN	State	3–8	✔		✔	✔
TX	State	3–9	✔		✔	✔
UT	State/ITBS	2–11/K–8	✔	✔	✔	✔
VT	NECAP	3–8	✔		✔	✔
VA	State	3–8	✔		✔	✔
WA	State	3–8	✔		✔	✔
WV	State, ACT	3–8	✔	✔	✔	✔
WI	State	3–11/8	✔		✔	✔
WY	State	3–8, 11	✔		✔	✔

Appendix B.
Alternative Measures Currently Used in Teacher Evaluation

Teacher Work Sample (TWS)–Oregon

What it is. Teacher work samples are widely used in a number of states, although their purpose and character vary substantially. Broadly speaking, teacher work samples are designed to demonstrate a teacher's ability to assess, plan, and implement effective instruction to students and can be used as both a pedagogical model for teacher education and a teacher assessment tool. Teacher work samples are employed in various ways in California, Colorado, Delaware, Georgia, Hawaii, Idaho, Iowa, Kansas, Kentucky, Louisiana, Michigan, Missouri, New York, North Carolina, Ohio, Oklahoma, Pennsylvania, Texas, Utah, Virginia, and Washington.

The teacher work sample in Oregon is foundational to teacher preparation in the state. Although successfully completing two work samples is a formal requirement for initial licensure in the state, only a handful of teacher candidates have been denied licensure on the basis of poor work sample performance because of the way it is embedded in teacher preparation. As a result, the TWS is more of a formative, pedagogical tool for pre-service teacher preparation than it is a summative, high-stakes assessment.

How it works. The Teacher Work Sample Methodology (TWSM) developed and defined by faculty at Western Oregon University is a "written, standards-based contextual teaching and learning unit that demonstrates a candidate's ability to assess, plan, and instruct in a standards-based educational system and impact student learning in a positive manner" (January 13, 2009, presentation to task force). In the state of Oregon, teacher candidates are required to successfully implement two teacher work samples prior to being awarded an initial teaching license and are encouraged (but not required) to complete additional work samples for attaining second-stage licensure. Teacher education programs in the state use the TWS data to assess an individual candidate's ability to teach to state and national standards, to enact best practices in content-based pedagogy linked to national professional standards, and to impact student learning. Aggregate

data from the teacher work samples are also used for program accountability, program improvement, and as a context for research.

How it links student learning to teacher performance. Two of the eight principles that guide the development and use of the teacher work sample methodology in Oregon place a clear emphasis on student learning:

- Judgments about a candidate's effectiveness as a teacher need to take into account the gains in learning made by every student taught.

- Documentation of a candidate's effectiveness as a teacher needs to be accompanied by observations of practice and descriptions of context, as well as evidence of learning gains by students.

The teacher work sample assesses a set of skills that facilitates the connection between teaching and learning and requires that teacher candidates develop specific products—or work samples—that demonstrate those skills. The full work sample involves the development of a unit of instruction, which includes at least 10 lessons. When a candidate successfully weaves together these skills into a comprehensive teacher work sample, the developers assert that the result stands as evidence that teaching and learning have been connected successfully.

A more detailed summary of the TWSM "underlying skills" and the ways in which candidates must demonstrate these skills is provided in *Appendix D*.

Renaissance Teacher Work Sample

What it is. Borrowing teacher work sample methodology concepts developed at Western Oregon University, members of the Renaissance Consortium (with leadership at Western Kentucky University), a consortium of 11 teacher preparation institutions from across the country, designed its teacher work sample around seven teaching processes it believed were critical to producing improved P–12 student learning. These are summarized in *Appendix C*.

How it works. As in Oregon, the teacher work samples developed by Renaissance member institutions (and housed at Western Kentucky University) are used in teacher preparation. Unlike Oregon, however, the work samples are not used as part of state certification decisions. Consortium members are currently engaged in a four- to five-year reliability study of inter-rater agreement on the work samples with the near-term goal of requiring candidates to earn at least a two on a three-point scale to receive a passing grade in student teaching by the end of 2009. Currently, researchers estimate that they get about 75–80 percent agreement on the overall score of the teacher work sample, and the scoring system is compensatory (that is, candidates can miss an entire dimension but still pass by making up ground on other components). Both of these factors suggest caution in using the measures for high-stakes decisions.

How it links student learning to teacher performance. The foundation of the Renaissance teacher work sample is a set of teaching practices deemed crucial to improving P–12 student learning, including the use of "pre-post" measures and formative assessment to guide instruction and the analysis and reporting of learning for all students and significant groups.

Denver Professional Compensation System for Teachers Program

What it is. The Denver "Pro-Comp" program is a performance-based teacher pay system that has been in effect since 2006 for members of the Denver Public Schools teachers' union, the Denver Classroom Teacher Association. Similar programs are being piloted or implemented in school districts in Austin, Texas; Helena, Montana; Charlotte-Mecklenburg, North Carolina; Catalina Foothills, Arizona; and Steamboat Springs, Colorado.

How it works. All new teachers hired in the Denver Public Schools system are automatically part of the program; teachers who were in the system when the program was implemented in 2006 could choose to opt in or remain in the existing salary scale.

The compensation system includes four main components: knowledge and skills; professional evaluation; market incentives; and student growth. The largest portion of the new funds used for Pro-Comp's compensation system is obligated to the knowledge and skills element, whose

purpose is to recognize and reward teachers who continue to develop and demonstrate skills and knowledge in their specific discipline. The professional evaluation system component is designed to recognize and reward teachers who demonstrate proficient practice through a professional evaluation. The market incentive component provides payments to teachers who accept positions in schools designated by the Denver Public Schools as hard-to-serve (for example, schools with large populations of students living in poverty) or hard-to-staff (for example, shortage areas in DPS such as middle-school mathematics teachers).

How it links student learning to teacher performance. Most relevant, perhaps, is that the final component of the Pro-Comp system is student growth, which is designed to reward teachers whose students meet and exceed expectations for academic growth. This component has three elements: instructional objective setting; Colorado Student Assessment Program (CSAP, the state test) incentive; and distinguished schools. The bulk of the money is allocated to the instructional objective-setting element, which involves a district-wide annual process in which each teacher, with his or her supervisor, sets two student growth objectives. If teachers participating in Pro-Comp meet both objectives, they earn a 1 percent (of an index) increase in their base salary; if they meet one objective, they earn a 1 percent bonus. The guidelines for Pro-Comp expressly forbid the use of CSAP measures in the assessment of the student growth objectives. An excerpt from a guidebook describing Pro-Comp explains it this way:

> Students whose teachers developed the highest quality objectives … average greater gains in achievement on the ITBS—whether the objectives were met or not met—than students whose teacher objectives were scored lower on the rubric. The same was true for CSAP scores …. [T]he better way to measure the impact of a teacher on the lives of students was through student growth measures and, better yet, multiple growth measures. Therefore, CSAP is not permitted to be used in writing student growth objectives.

In other words, the program rewards objective-setting as a process, because its developers view it as a good instructional practice that contributes to student learning by focusing instruction. Elementary teachers are expected to write one objective in reading and one in mathematics, and

secondary teachers write objectives according to the subject they teach. After objective-setting in the fall, progress is assessed mid-year, and adjustments are made as necessary. In the spring, each teacher's supervisor assesses how many objectives have been met. The items in the checklist that must be used in developing each objective are described in *Appendix F*.

Finally, Pro-Comp includes criteria for the assessments that are used to measure student growth in the objectives. The assessments must measure the learning content of the objective and be closely tied to the curriculum and, when available, are to rely on district-approved assessments that reflect what students are expected to learn in the courses they teach. Thus, the assessments used are a mix of district-developed assessments, commercially available measurement tools, and assessments developed by individual teachers to measure progress toward individual objectives.

The second element of the student growth system component is the CSAP incentive element, which ties student performance on the state test to teacher pay. This 3 percent salary increase is awarded to teachers whose students significantly exceed the expected range of improvement for one year's growth. These increases continue as long as the teacher's students continue to exceed the expected growth pattern; if the students fall below the lower limit of the standard range, the teacher loses the increase. The program estimates that at most 30 percent of teachers would be eligible for such pay. Because the element is based on growth from the previous year, it is available only to teachers of mathematics and language arts in grades 4 through 10.

Finally, the distinguished schools element of the student growth system component is a bonus for serving in a distinguished school, based on multiple measures of school quality.

Arizona Career Ladder

What it is. The Arizona Career Ladder Program, like the Denver Pro-Comp, uses measures of student learning as part of its performance-based compensation plan that provides incentives to teachers in 28 Arizona school districts who choose to make career advancements without leaving the classroom or the profession.

How it works. The Career Ladder requires evaluating and compensating teachers based on their level of skill attainment and demonstrated student academic progress, rather than as a result of seniority and educational credits.

How it links student learning to teacher performance. The measures of student learning used in the program are determined locally. Some are locally designed, others are state or national standards-based or norm-referenced assessments, and some are diagnostic or prescriptive assessments. They are reviewed by the State Career Ladder Advisory Committee and ultimately approved by the State Board of Education. As a result, the measures vary across and even within the 28 jurisdictions participating in the state program.

Teachers prepare a Career Ladder portfolio that includes evidence to meet the legislative requirement that they be able to gauge "increasingly higher levels of pupil academic progress as measured by objective criteria." The portfolio, gathered over the course of a given school year, includes three components: (1) Evaluation of Teacher Performance; (2) Evaluation of Teacher's Pupil Academic Progress; and (3) Professional Development and Higher Level Instructional Responsibilities. Teachers work independently or in groups or teams to complete their portfolios.

Two levels of pupil progress are submitted by the individual teacher or team and by the school and district as a whole. The district data are submitted each year to the Career Ladder Advisory Committee (CLAC) as a part of its annual evaluation and application for continuation in the program; funding is conditional upon completing this requirement. These data provide a summary of any pertinent district assessments used by the district—for example, Dynamic Indicators of Basic Early Literacy Skills (DIBELS)—as well as a summary of the Arizona Instrument to Measure Students (AIMS) and TerraNova (CBT McGraw-Hill's standardized assessment). Some districts also use the Arizona English Language Learner Assessment (AZELLA) or other language proficiency assessments. At that level, the CLAC is looking for overall evidence of gain in areas that meet the goals set out by the district and the Career Ladder Program. CLAC looks for a match in professional development focus and alignment with program goals and objectives, as well as trends in overall student growth and achievement.

As a part of its yearly application, each Career Ladder District submits to the CLAC a student progress or student achievement plan template or description that includes the requirements for that district. Teachers who participate in the Career Ladder program must complete student assessment plans as a part of their portfolio. The purpose of the student progress component is twofold: (1) to focus a teacher's attention on increasing student achievement at the classroom level in a particular set of skills that his or her students need to improve, and (2) to demonstrate the overall effectiveness of reflective practice and targeted professional development on student achievement from a site or district perspective for the Career Ladder Program.

Teachers generally choose one subject area that is consistent with overall school goals for improvement, such as reading or math. These plans require an assessment of current student achievement levels (pre-tests or analysis of current data); defined goals and objectives for instruction (aligned with the state standards); evidence of formative assessments; and a summative assessment of progress as well as an analysis of the data and instructional factors that may have affected the results. Teachers in the very early grades or those who teach special areas often use teacher-made or curriculum-based assessments for their classroom pre-tests and formative assessments and use some form of curriculum-based, district, or state assessment for the summative, or overall progress. In the upper grades, where more longitudinal data are available, they sometimes use a particular portion of the state assessment, or content assessments that are a part of their curriculum. Some of the districts have begun using online assessment tools that are compatible with particular content areas that they have purchased to complement their curriculum. Each teacher's portfolio is individually scored by his or her peers at the district level based upon rubrics that are developed. A teacher's placement on the Career Ladder and the financial addendum he or she receives depends on the results of this local evaluation. ★

Appendix C.
Experimental Instruments to Assess Teaching Practice

Along with these integrated approaches developed by selected districts and states, the task force identified a series of experimental instruments that more specifically assess teaching practice. While some of these instruments are still being refined to address shortcomings, they can serve as a source of ideas about expanded or alternate methods of incorporating direct measures of teaching practice in a classroom setting into broader evaluations of teacher effectiveness. They include:

The Classroom Assessment Scoring System (CLASS)

What it is. A standardized observation mechanism that focuses on teacher-student interaction in early childhood and elementary classrooms.

How it works. CLASS categorizes effective teacher-student interactions in preK–3 classrooms into three broad domains: emotional support, classroom organization, and instructional support. The program's developers point to research from 3,000 classrooms suggesting that improvements in effective interactions in the three areas measured by CLASS translate into improved achievement and social skill development in young children. The program's developers are creating tools to facilitate the system's use in teacher preparation and education, professional development, program monitoring, and research and evaluation.

The Learning Mathematics for Teaching (LMT) Project

What it is. A coding rubric that measures the quality of mathematics instruction by evaluating teachers' understanding of and ability to apply mathematical knowledge in the classroom.

How it works. LMT focuses on identifying the mathematical knowledge needed for effective teaching. Its coding rubric focuses on such domains as the teacher's ability to work with "rich mathematics," meaning the concepts behind computation; the presence of mathematical errors and imprecise language in instruction; connecting classroom work to mathematical concepts; checking for student understanding; and the cognitive level of student work.

The Protocol for Language Arts Teaching Observations (PLATO)

What it is. An observation system focusing on 10 dimensions of instruction in English and language arts classrooms.

How it works. Designed for middle and high school English/language arts classrooms, PLATO incorporates classroom organization and emotional support elements from the CLASS domains, as well as content domains that cut across ELA subject areas, including reading, writing, literature, speaking and listening, and grammar and mechanics. PLATO examines 10 elements of instruction: purpose, intellectual challenge, representations of content, connections to personal or prior knowledge, models and modeling, explicit strategy instruction, guided practice, feedback, classroom discourse, and ELL accommodations. ★

Appendix D.
Summary of Underlying Skills and Their Demonstration in the Western Oregon University Teacher Work Sample Methodology

Analysis of context. Teacher candidates must identify and analyze the contextual factors that shape teaching and learning. This skill is gauged by assessing the candidate's description of the educational setting and discussion of the potential effects of teaching and learning.

Selection of content. Teacher candidates must select important, powerful, developmentally appropriate, and useful content that appeals to students and their surrounding community and reflects state and national standards. This skill is gauged by assessing the candidate's stated goals and objectives and his or her depiction of how the selected content aligns with professional standards.

Selection of pedagogy. Teacher candidates must select pedagogical strategies that are aligned with context, content, and prior student knowledge. This skill is gauged by assessing the candidate's selection and justification of pedagogical strategies in his or her lesson plans.

Use of assessment. Teacher candidates must design measures and collect data before, during, and after instruction. This skill is gauged by assessing the candidate's assessment plan that specifies pre-, post-, and formative measures; articulates their validity and reliability; and connects them to the goals and objectives of the instructional unit.

Data analysis. Teacher candidates must analyze aggregated and disaggregated data before, during, and after instruction. This skill is gauged by assessing the candidate's reports on student learning at the individual student level and in various groupings.

Reflective analysis. Teacher candidates must reflect on their work; the progress and engagement of their students; and the interaction and alignment among setting, content, pedagogy, and assessment. This skill is gauged by assessing the candidate's essay, developed after the teacher work samples exercises are complete, which reflects on his or her effectiveness in helping all students reach the defined goals and objectives.

Alignment. Teacher candidates must align assessment procedures, learning experiences, goals and objectives, and contextual factors. This skill is gauged by a holistic evaluation of the teacher's work sample products. ★

Appendix E.
Teaching Practices Deemed Crucial to Producing Learning in P–12 Students by the Renaissance Teacher Work Sample Methodology

::::::: Use of student and classroom context to design instruction

::::::: Use of instructional unit learning goals that addressed local and state content standards

::::::: Use of pre-, post-, and formative assessment to guide instruction and measure and report learning results

::::::: Design of instruction for all students that addressed unit learning goals and was aligned with concepts and processes assessed

::::::: Instructional decision making based on continuous formative assessment

::::::: Analysis and reporting of learning for all students and significant groups

::::::: Reflection and evaluation of teaching and learning

Appendix F.
Denver Pro-Comp Program Checklist for Developing Student Learning Objectives

☐ Rationale—why that particular objective was chosen

☐ Population—which students the objective addresses

☐ Interval of time—weeks, quarters, semesters, school year

☐ Assessment used to measure whether the objective was met (pre- and post-data)

☐ Expected gain or growth made by the students (the heart of the objective)

☐ Learning content—the academic skills, behavior, or attitudes teachers are trying to support, based on needs identified in the baseline data; includes realistic personal goal-setting and problem-solving strategies

☐ Strategies—teaching methods or interventions by service professionals to be used to achieve the objective; include one-on-one contact, home visits, referral to extra-curricular activities

NBPTS®

National Board for
Professional Teaching Standards

1525 Wilson Boulevard, Suite 500
Arlington, VA, 22209
1-800-22TEACH
www.nbpts.org