

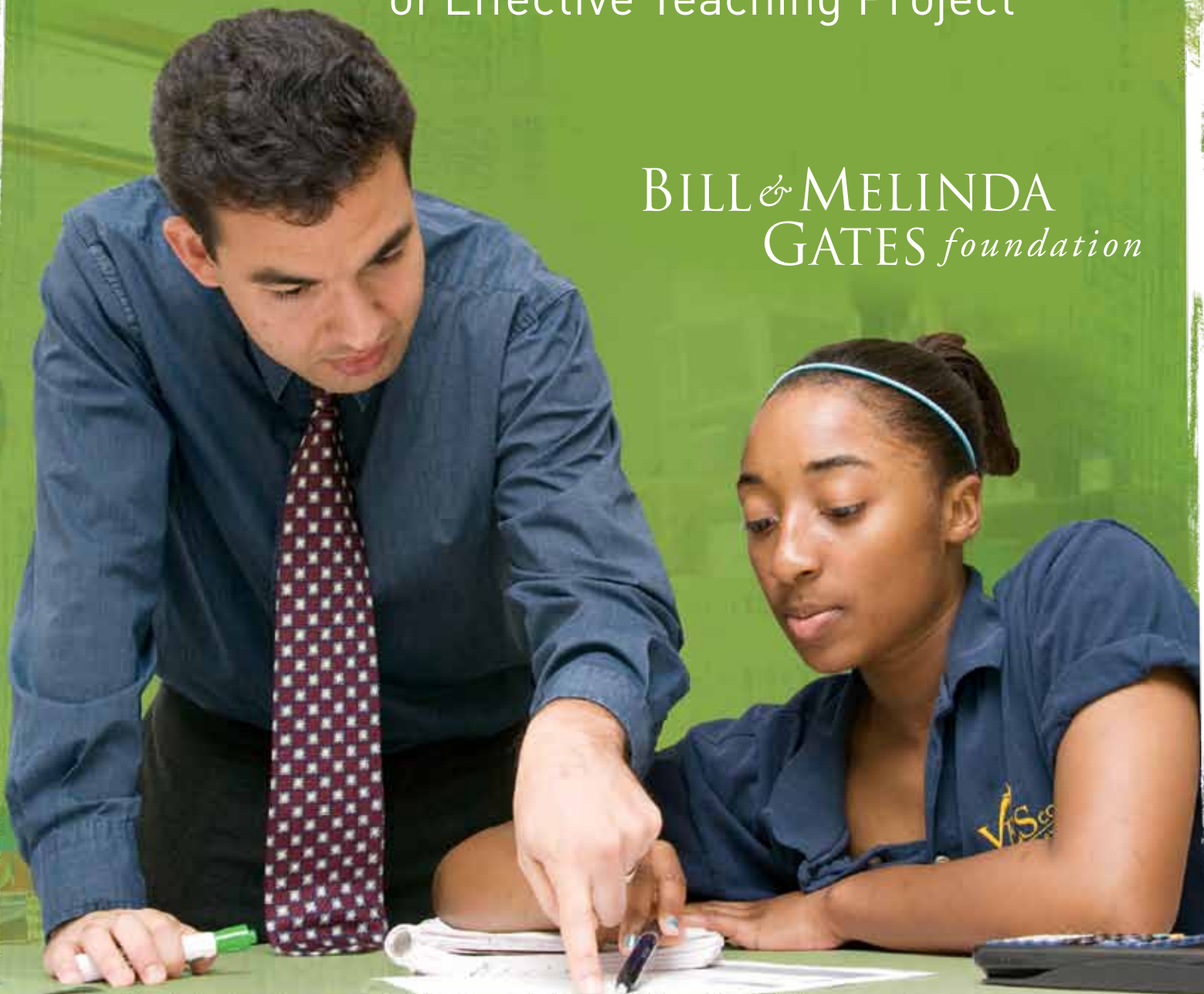
MET
project

POLICY BRIEF

Learning about Teaching

Initial Findings from the Measures
of Effective Teaching Project

BILL & MELINDA
GATES *foundation*



Introduction

For four decades, education research has confirmed what many parents know: A child's learning depends on the talent and skills of the person leading his or her classroom. As much as parents worry about their local school, most eventually learn that their child's teacher in that school matters even more.



Yet most school systems ignore differences among individual teachers. Information about teaching effectiveness is neither collected nor shared. The costs of this neglect are enormous. Novice teachers' skills plateau far too early without the feedback they need to improve. Likewise, there are too few opportunities for experienced teachers to share their practice and strengthen the profession. Finally, principals are forced to make the most important decision we ask of them—granting tenure to beginning teachers still early in their careers—with little objective information to guide them.

If we say “teachers matter” (and the research clearly says they do), why do we pay so little attention to the work teachers do in the classroom? If teachers are producing dramatically different results, why don't we provide them with that feedback and trust them to respond?

The MET Project

In fall 2009, the Bill & Melinda Gates Foundation launched the Measures of Effective Teaching (MET) project to test new approaches to recognizing effective teaching. Our goal is to help build fair and reliable systems for teacher observation and feedback to help teachers improve and administrators make better personnel decisions. To be sure, great teaching has many intangible qualities. However, we set out to test whether there are aspects of effective teaching—such as effectively managing a classroom, starting each class with a clear objective, engaging students with questioning strategies, consolidating the lesson at the end of a period, and diagnosing common student errors and correcting them—that can be systematically measured by observing

The MET Project

In fall 2009, the Bill & Melinda Gates Foundation launched the Measures of Effective Teaching (MET) project to test new approaches to recognizing effective teaching. The project's goal is to help build fair and reliable systems for teacher observation and feedback to help teachers improve and administrators make better personnel decisions. With funding from the foundation, the data collection and analysis are being led by researchers from academic institutions, nonprofit organizations, and several private firms and are being carried out in seven urban school districts.

Research Partners

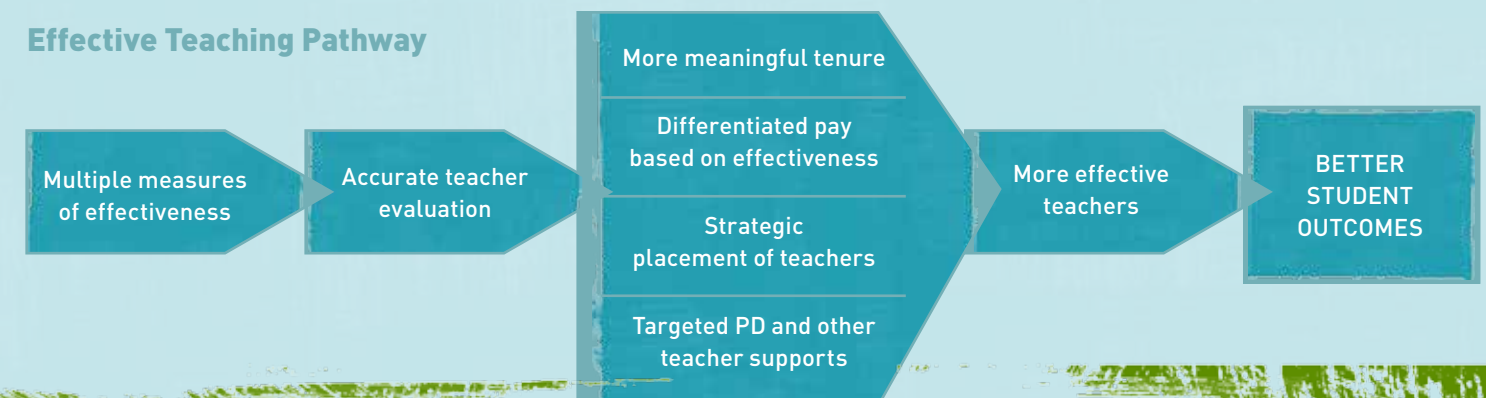
Our lead research partners include:

- Mark Atkinson, Teachscape
- Nancy Caldwell, Westat
- Ron Ferguson, Harvard University
- Drew Gitomer, Educational Testing Service
- Eric Hirsch, New Teacher Center
- Dan McCaffrey, RAND
- Roy Pea, Stanford University
- Geoffrey Phelps, Educational Testing Service
- Rob Ramsdell, Cambridge Education
- Doug Staiger, Dartmouth College

Key contributors include:

- Joan Aucter, National Board for Professional Teaching Standards
- Charlotte Danielson, The Danielson Group
- Dan Goldhaber, University of Washington
- Pam Grossman, Stanford University
- Bridget Hamre, University of Virginia
- Heather Hill, Harvard University
- Sabrina Laine, American Institutes for Research
- Catherine McClellan, Educational Testing Service
- Denis Newman, Empirical Education
- Raymond Pecheone, Stanford University
- Robert Pianta, University of Virginia
- Morgan Polikoff, University of Southern California
- Steve Raudenbush, University of Chicago
- John Winn, National Math and Science Initiative

Effective Teaching Pathway



Our Analysis

classrooms and by asking students. If so, such measures would be useful for both developing teachers and staffing schools more effectively.

With funding from the foundation, the data collection and analysis are being led by researchers from academic institutions, nonprofit organizations, and several private firms. However, the hardest work is being done by the 3,000 teacher-volunteers, working in seven urban school districts (New York City, Charlotte-Mecklenburg, Hillsborough County in Florida, Memphis, Dallas, Denver, and Pittsburgh¹), who have agreed to open their classrooms.

Although the project is ongoing (the final report will not be released until winter 2011–12), we are reporting our findings as they become available in order to inform the important reform work already under way in states and districts around the country. This is the first such report.

Data Collection So Far

Last spring, we collected **digital video for 13,000 lessons** in the classrooms taught by our teacher-volunteers. Eventually, we will score each of those lessons using several protocols (or rubrics) that may help identify effective teaching in the classroom. There are literally thousands of interactions between a teacher and students every day. We will want to know which aspects of instruction are most strongly related to student achievement gains so that supervisors can focus their feedback on the things that matter most.

We also asked **students to report their perceptions of each teacher's classroom**.² We wanted to know if students' perceptions of the learning environment in a teacher's classroom are consistent with the learning gains they experience. In addition, we asked students to take an **assessment to supplement their scores on the state test**. Students in grades 4 through 8 math classes were assessed for their conceptual understanding of

key concepts in mathematics (using the Balanced Assessment of Mathematics), while students in English language arts classes were asked to read short passages and provide written responses to questions probing their comprehension (using the open-ended version of the Stanford Achievement Test, 9th Edition for reading). We tested high school students using the Quality Core end-of-course assessments from ACT, in Algebra I, 9th grade English, and Biology.

For this report, we have studied student achievement gains on the state test and the supplemental tests in grades 4 through 8 for five MET districts.³ We also have studied student perception data in these 4th to 8th grade classrooms. However, because we have scored only a fraction (roughly 6 percent) of the lesson videos using only two of the assessments of classroom practice, it is too early to conclude which approaches to classroom observation are most helpful or which aspects of such observations are most telling.

As a school leader staffs a school each year, he or she must ask, “*What does each teacher's past performance say about his or her ability to help students learn?*” and “*What are his or her specific strengths and weaknesses?*” Every artifact of a teacher's practice—whether student surveys about a teacher's effectiveness, direct classroom observations, or (in an increasing number of school districts) the achievement gains of recent or past students—is potentially useful in identifying a teacher's strengths and weaknesses and prospects of success with future students. Effective leaders can use such data to guide a teacher's development.

Our analysis plan mimics the school leader's questions. We ask, “*How well do various aspects of a teacher's performance in one course section or in one academic year help predict student achievement gains in that teacher's classroom during another academic year or in another course section?*” In this preliminary report, we measure student achievement gains using two different tests in each subject, the state standardized test and an additional, more cognitively demanding test. In the future, we anticipate expanding these outcomes beyond traditional tests to include noncognitive measures as well. For now, we test the value of evidence of effectiveness from one class in anticipating student achievement gains in another class taught by the same teacher. To

do that, we use two analogous thought experiments:

- First, focusing on the subset of teachers for whom we have measures from *more than one* classroom of students during 2009–10, we ask whether the measures of practice from one class predict the teacher's contribution to student learning gains in *another* class.
- Second, focusing on those teachers for whom we have student assessment data from a prior year (2008–09), we test whether measures of classroom practice in 2009–10 are related to the teacher's contribution to student learning gains in another school year.



1 Pittsburgh served as our pilot district, an important role, but no data from this district will be analyzed.

2 The Tripod survey, which we used, was developed over the past decade by Dr. Ron Ferguson from Harvard in collaboration with Cambridge Education.

3 The results from Memphis have been delayed because of a new state test in Tennessee last spring. Moreover, we are still organizing the data for the high school students from the other districts.

If the measures are accurate in predicting performance in other school years and in other classes, they will help teachers focus on the areas of their practice that need developing and help principals make more discerning personnel decisions.

Early Findings

Although the accompanying technical report provides many more details on our analysis and initial results, we have four general findings to report:

First, in every grade and subject we studied, a teacher’s past success in raising student achievement on state tests (that is, his or her *value-added*) is one of the strongest predictors of his or her ability to do so again.

When applied to teaching, the term *value-added* refers to statistical efforts to isolate the *impact* of a teacher on his or her students’ achievement by

adjusting for each student’s starting point coming into the class. Each student’s performance at the end of the year is then compared to that of similar students elsewhere (with similar prior test scores, similar demographics, etc.). When a teacher’s students outperform his or her peers whose students have similar prior achievement, characteristics, and classmates, it constitutes positive student growth or value-added. (In this analysis, we also adjusted for the mean characteristics of the other students in the class, since one’s peers also can have an influence on one’s learning.) Conversely, when a teacher’s students perform worse than his or her peers whose students have similar starting points and similar classmates, it constitutes negative growth or value-added.

A teacher’s history of positive (or negative) value-added is among the strongest predictors of his or her students’ achievement growth in other classes

and academic years. Value-added methods have been criticized as being too imprecise, since they depend on the performance of a limited number of students in each classroom. Indeed, we do find that a teacher’s value-added fluctuates from year to year and from class to class, as succeeding cohorts of students move through his or her classrooms. However, our analysis shows that volatility is not so large as to undercut the usefulness of value-added as an indicator of future performance.

Second, the teachers with the highest value-added scores on state tests *also* tend to help students understand math concepts or demonstrate reading comprehension through writing.

Many have speculated that teachers with high value-added scores are simply coaching students to score well on the state tests. If this were true, value-added data would be of limited value in identifying effective teaching—even if they were predictive. After all, it would do students little good to score well on state tests if they failed to understand key concepts. We don’t see that. Rather, we see evidence that teachers with high value-added on state tests also seem to help students perform better on the supplemental tests. This seems particularly true in mathematics.

Some of the classrooms in our study did focus on test preparation. In many classrooms students reported that

Student Perceptions Matter

The 7 Cs	Sample Questions	Percentage of Students Agreeing with Each Item	
		At the 25th percentile	At the 75th percentile
CARE	My teacher in this class makes me feel that s/he really cares about me.	40%	73%
	My teacher really tries to understand how students feel about things.	35%	68%
CONTROL	Students in this class treat the teacher with respect.	33%	79%
	Our class stays busy and doesn’t waste time.	36%	69%
CLARIFY	My teacher has several good ways to explain each topic that we cover in this class.	53%	82%
	My teacher explains difficult things clearly.	50%	79%
CHALLENGE	In this class, we learn a lot almost every day.	52%	81%
	In this class, we learn to correct our mistakes.	56%	83%
CAPTIVATE	My teacher makes lessons interesting.	33%	70%
	I like the ways we learn in this class.	47%	81%
CONFER	Students speak up and share their ideas about class work.	40%	68%
	My teacher respects my ideas and suggestions.	46%	75%
CONSOLIDATE	My teacher checks to make sure we understand what s/he is teaching us.	58%	86%
	The comments that I get on my work in this class help me understand how to improve.	46%	74%

Survey items are differentiated based on grade level and can be administered online or on paper.

The table above, based on the Tripod survey, shows that students are able to differentiate between teachers and their classroom environments. The Tripod survey identifies seven constructs—the 7 Cs—that are core to a student’s experience in his or her classroom. For example, “Care” refers to the extent to which students report that their teacher cares about them as measured by multiple survey questions. “Control” refers to the extent to which teachers effectively manage student behavior in the classroom.

“We spend a lot of time in this class practicing for the state test,” or “Getting ready for the state test takes a lot of time in our class.” *However, the teachers in such classrooms rarely show the highest value-added on state tests.* On the contrary, the type of teaching that leads to gains on the state tests corresponds with better performance on cognitively challenging tasks and tasks that require deeper conceptual understanding, such as writing.

Third, the average student knows effective teaching when he or she experiences it.

When a teacher teaches multiple classes, student perceptions of his or her practice are remarkably consistent across

different groups of students. Moreover, student perceptions in one class or one academic year predict large differences in student achievement gains in other classes taught by the same teacher, especially in math. In other words, when students report positive classroom experiences, those classrooms tend to achieve greater learning gains, and other classrooms taught by the same teacher appear to do so as well.

Student feedback need not be a popularity contest. We asked detailed questions about various aspects of students’ experience in a given teacher’s classroom. Some questions had a stronger relationship to a teacher’s value-added than others. The most predictive aspects

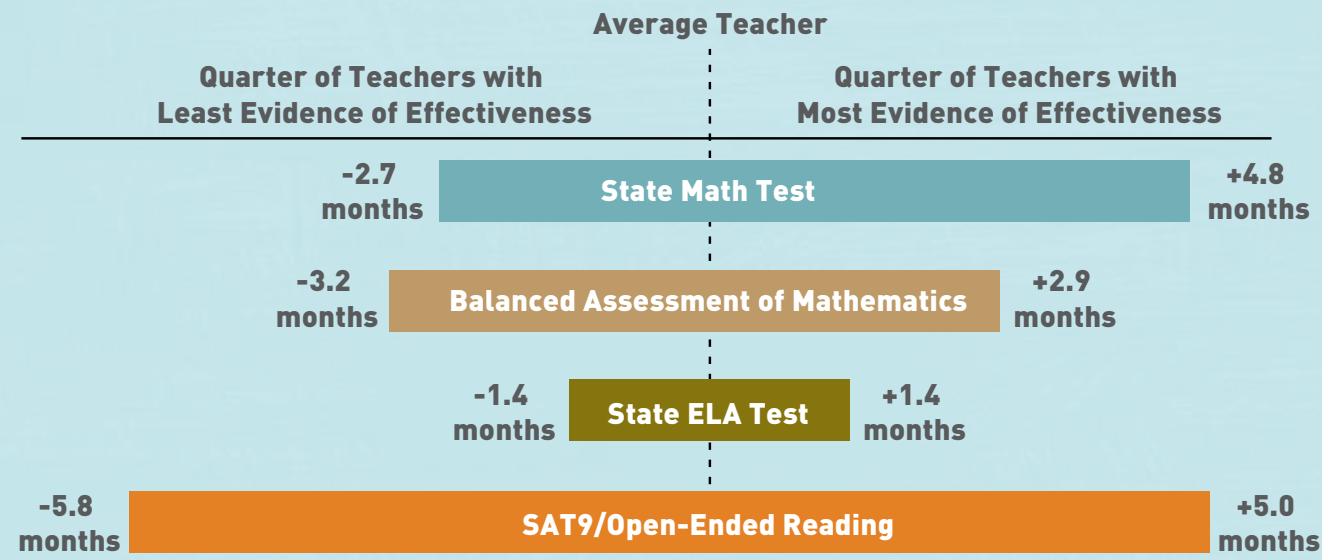
of student perceptions are related to a teacher’s ability to control a classroom and to challenge students with rigorous work.

Students’ perceptions have two other welcome characteristics: They provide a potentially important measure that can be used in nontested grades and subjects. In addition, the information received by the teacher is more specific and actionable than value-added scores or test results alone.

Fourth, valid feedback need not be limited to test scores alone. By combining different sources of data, it is possible to provide diagnostic, targeted feedback to teachers who are eager to improve.



Students with Most Effective Teachers Learn More in a School Year



Months of learning gain are calculated based on the difference in value-added gains between the top and bottom quartile of teachers compared to the average teacher. The number of months of schooling applies to a nine-month school year, using a .25 standard deviation per year conversion factor.

The public discussion usually portrays only two options: the status quo (where there is no meaningful feedback for teachers) and a seemingly extreme world in which tests scores alone determine a teacher's fate. Our results suggest that's a false choice. It is possible to combine measures from different sources to get a more complete picture of teaching practice. The measures should allow a school leader to both discern a teacher's ability to produce results and offer specific diagnostic feedback. Value-added scores alone, while important, do not recommend specific ways for teachers to improve.

Ultimately, we will be adding data from classroom observations and a new teacher assessment to the mix of measures we are testing. However, our initial analyses suggest that the combined

measures help identify effective and ineffective teaching. For example, we used evidence of a teacher's performance (as measured by value-added and student perceptions) in one class to infer which teachers were more and less effective. We then assessed the impact of these teachers on learning

gains for a *different* group of students. As shown by the "State Math Test" bar in the graphic above, students of math teachers whose value-added scores and student perceptions placed them in the bottom 25 percent gained the equivalent of only six and a half months of learning during a nine-month school year as

Percentage of students who agreed with the following statements

	Bottom 25% of effective teachers	Top 25% of effective teachers
Our class stays busy and does not waste time.	38%	64%
My teacher explains difficult things clearly.	48%	76%
I like the ways we learn in this class.	49%	77%
We learn a lot in this class every day.	56%	79%

First we sorted teachers based on student perception surveys and value-added on the state math assessment. Then we sorted teachers into quartiles. The percentage of students agreeing above represents the mean for the top and the bottom quartile teachers.

measured by the state math assessment. Their students were clearly shortchanged. However, students of those math teachers identified to be in the top 25 percent gained nearly 14 months of learning during this same nine-month school year. The difference in learning associated with being assigned a top quartile teacher rather than a bottom quartile teacher was more than seven months—roughly two-thirds of a school year!

Given these large differences, it is vital that we identify specific areas of practice where struggling teachers could improve—such as managing class time more effectively. More examples are in the table at the bottom of page 6.

While the student survey data are quite encouraging, we expect the additional information provided by the other measures, such as the classroom observation protocols and the teacher

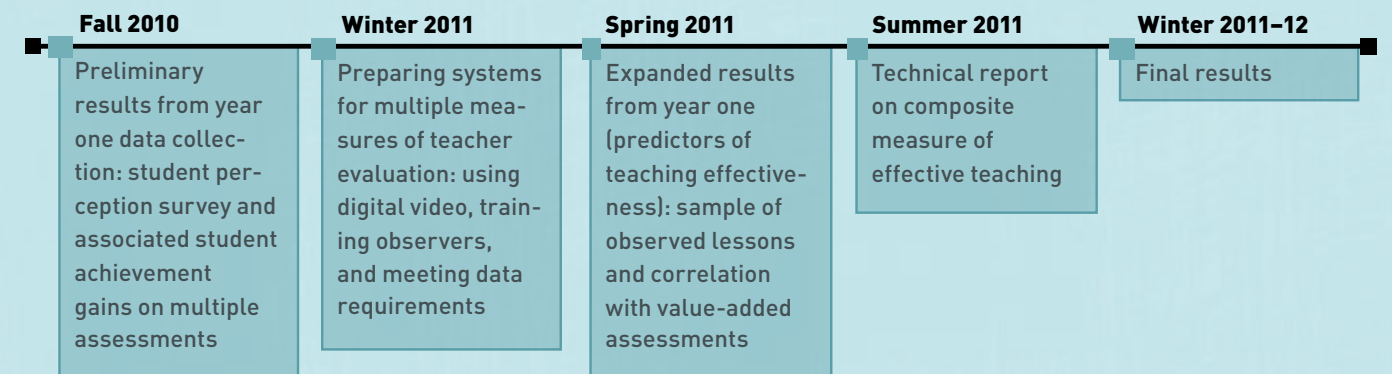
knowledge assessment, to yield even greater insights into the different knowledge, skills, and practices adopted by the most and least effective teachers.

Still to Come

As noted above, we're far from done with the MET project. We still need to complete the analysis of 13,000 classroom lessons observed during the 2009–10 school year and the fresh set of lessons from the current school year. We also will test a new measure that extends and refines the concept of pedagogical content knowledge for teachers, or what a teacher knows about *how* to teach a particular subject. These findings could have significant implications, not only for measuring effective teaching but for the training and development of teachers as well.

In late spring 2011, we will issue a more complete report from year one that includes findings from the classroom observation protocols. Late summer of 2011 researchers from RAND will combine data from each of the MET project measures to form a "composite indicator" of effective teaching. Researchers from RAND will analyze different approaches to weighting each measure (student achievement on state and supplemental tests, classroom observations, teacher knowledge, student perceptions) when forming an overall assessment of a teacher's effectiveness. Finally, early in 2012, we will report whether those teachers whose performance was rated most highly during the 2009–10 school year actually produced larger student achievement gains than their colleagues during the 2010–11 school year.

MET Project Implementation Timeline



Conclusion

Reinventing the way we develop and evaluate teachers will require a thorough culture change in our schools. No longer should teachers expect to close the door to their classrooms and “go it alone.” The quality of instruction is a public good, and improvement will require a collective commitment to excellence in every classroom.

Teachers will need to open up their practice for review and constructive critique—because that’s what excellence requires.

There are some obvious places to start now:

- working with teachers to develop accurate lists of the students in their care, so that value-added data are as accurate as possible
- using confidential surveys to collect student feedback on specific aspects of a teacher’s practice, including those in nontested grades and subjects
- retraining those who do classroom observations to provide more meaningful feedback

While we still have much to learn in the future, we can now confidently encourage states and districts to regularly check that the collection of measures they assemble allows them to explain the variation in student achievement gains among teachers. Just as we have done in the accompanying report, they should confirm that the measures as *implemented* continue to point in the

same direction. Even a great classroom observation tool can be implemented poorly (if principals are poorly trained or if they are unwilling to provide honest feedback). Even a great instrument for collecting student feedback can be distorted (if students do not take it seriously or if students do not trust that their answers will be kept confidential). The best way to ensure that the evaluation system is providing valid and reliable feedback to teachers is to regularly verify that—on average—those who shine in their evaluations are producing larger student achievement gains.

Since we are just starting, we need to be humble about what we know and do not know. However, we should take heart in the fact that the solutions to our educational challenges are implemented every day by those teachers who regularly generate impressive results. We just need to assemble the evidence on student achievement, ask students to help by providing their own confidential feedback, and refine our approach to classroom observation—to find those teachers who truly excel, support them, and develop others to generate similar results. The MET project is an important first step.

Bill & Melinda Gates Foundation

Guided by the belief that every life has equal value, the Bill & Melinda Gates Foundation works to help all people lead healthy, productive lives. In developing countries, it focuses on improving people’s health and giving them the chance to lift themselves out of hunger and extreme poverty. In the United States, it seeks to ensure that all people—especially those with the fewest resources—have access to the opportunities they need to succeed in school and life. Based in Seattle, Washington, the foundation is led by CEO Jeff Raikes and Co-chair William H. Gates Sr., under the direction of Bill and Melinda Gates and Warren Buffett.

For more information on the U.S. Program, which works primarily to improve high school and postsecondary education, please visit www.gatesfoundation.org.

©2010 Bill & Melinda Gates Foundation. All Rights Reserved. Bill & Melinda Gates Foundation is a registered trademark in the United States and other countries.

BILL & MELINDA
GATES *foundation*

www.gatesfoundation.org