MET
project

Feedback for Better Teaching

Nine Principles for Using Measures of Effective Teaching

BILL & MELINDA GATES *foundation*

ABOUT THIS DOCUMENT: This brief highlights a set of guiding principles from the Bill & Melinda Gates Foundation to inform the design and implementation of high-quality teacher support and evaluation systems based on three years' of work by the Measures of Effective Teaching (MET) project, its partners, and other leading school systems and organizations. More in-depth discussion of the MET project's analyses to date may be found in the project's research reports and non-technical briefs at www.metproject.org.

ABOUT THE MET PROJECT: The MET project is a research partnership of academics, teachers, and education organizations committed to investigating better ways to identify and develop effective teaching. Funding is provided by the Bill & Melinda Gates Foundation.

The approximately 3,000 MET project teachers who volunteered to open up their classrooms for this work are from the following districts: The Charlotte-Mecklenburg Schools, the Dallas Independent Schools, the Denver Public Schools, the Hillsborough County Public Schools, the Memphis Public Schools, the New York City Schools, and the Pittsburgh Public Schools.

Partners include representatives of the following institutions and organizations: American Institutes for Research, Cambridge Education, University of Chicago, The Danielson Group, Dartmouth College, Educational Testing Service, Empirical Education, Harvard University, National Board for Professional Teaching Standards, National Math and Science Initiative, New Teacher Center, University of Michigan, RAND, Rutgers University, University of Southern California, Stanford University, Teachscape, University of Texas, University of Virginia, University of Washington, and Westat.

January 2013

Creating the Conditions for Success

Teachers want to succeed, but they typically lack the conditions for success. Teachers generally work in isolation. They're on their own to adjust practice to better serve students. What guidance they get often is plagued by vague teaching standards, overly numerous and often trivial learning objectives, and testing systems that measure only some of the outcomes that educators value for students. Success itself remains ill defined.

The partners in the Measures of Effective Teaching (MET) project—a group of thoughtful individuals in districts, unions, schools, research institutions, and technical organizations—told us from the outset that current evaluation systems were not being used to improve teacher support.

Nearly all of the teachers on the MET project's advisory panel similarly expressed little faith that traditional evaluation measures and practices could provide usable information to guide more effective teaching. They decried evaluation as perfunctory, its measures as disconnected from what they valued

about teaching and learning, and its observation practices as highly subjective. These advisors nevertheless agreed that trustworthy measures could inform improvements in teaching practice in ways that traditional evaluation systems have not. Identifying and validating better measures has been the primary goal of the MET project and a core concern of the districts with which we work.

Measuring for Improvement

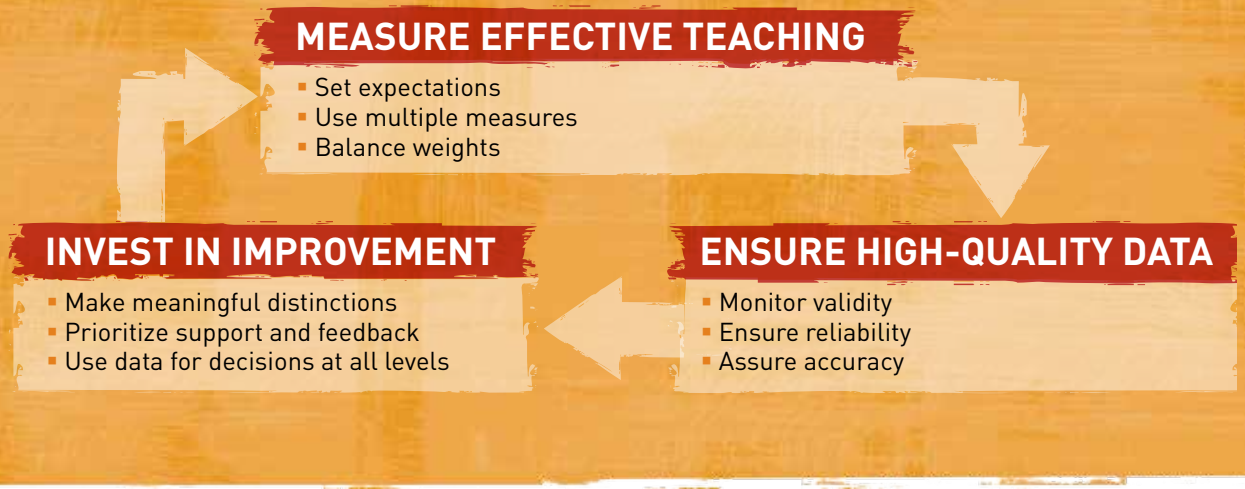
It is very hard to support effective teaching without good information about actual teaching practice. The MET project has sought to build and test measures of effective teaching so that school systems can clearly understand and then close the gap between their expectations for effective teaching and the actual teaching occurring in classrooms.

But good information is hard to produce. It requires the right measures, the right measurement processes, strong communications, and an awareness of how information can be distorted. When given the right type of attention, measures can help set expectations and align effort.

It will require care and attention for teacher evaluation measures to serve both professional development and accountability purposes. To help states and districts navigate the work of implementing feedback and evaluation systems that support teachers, we offer nine guiding principles based on three years' of study, observation, and collaboration with districts. Our prior reports tested, and ultimately supported, the claim that measures of teaching effectiveness could be valid and reliable. These principles, explained on the following pages, fall into three overarching imperatives, as shown in **Figure 1:** Measure Effective Teaching; Ensure High-Quality Data; and Invest in Improvement. Note the cyclical presentation. Well-designed evaluation systems will continually improve over time.

Figure 1

A Framework for Improvement-Focused Teacher Evaluation Systems



Guiding Principles for Improvement-Focused Teacher Evaluation Systems

Our district partners are beginning to build and implement systems for teacher feedback and evaluation. In each case, they have emphasized the importance of investing in improvement; they have upheld high standards for data quality; and they have included multiple measures. They see feedback as the path to better teaching. They understand that the measures, while focused on teaching, are able to provide feedback at all levels of the system—school leadership, coaching support, professional development, and even central office administration—to align efforts in support of more effective teaching and learning.

Measure Effective Teaching

- **Set Expectations.** The first step in designing teacher evaluation systems is for stakeholders to agree on the teacher knowledge, skills, and behaviors that enable better student learning. This benefits the entire system by providing a shared language to talk about teaching, set priorities, and target support. In the MET project, we defined effective teaching as sensitivity to students' academic and social needs; knowledge of subject-matter content and pedagogy; and the ability to put that knowledge into practice, all in the service of student success.

- **Use Multiple Measures.** The choice of measures should reflect the multifaceted nature of effective teaching. For instance, the MET project sought or developed measures to reflect all key aspects of its definition of effective teaching: student surveys to assess the supportiveness of the instructional environment; content tests to assess teachers' knowledge of their subject and how to teach it; observation instruments to assess teachers' classroom practice; and student assessments to measure the learning gains of a teacher's students. It was important that we measured each facet of effective teaching. An unmeasured facet is likely to be neglected.

- **Balance Weights.** What counts most gets the most attention. When combining measures into a single index, we have found that approaches that allocate between 33 percent and 50 percent of the weight to student achievement measures are sufficient to indicate meaningful differences

continues on page 6

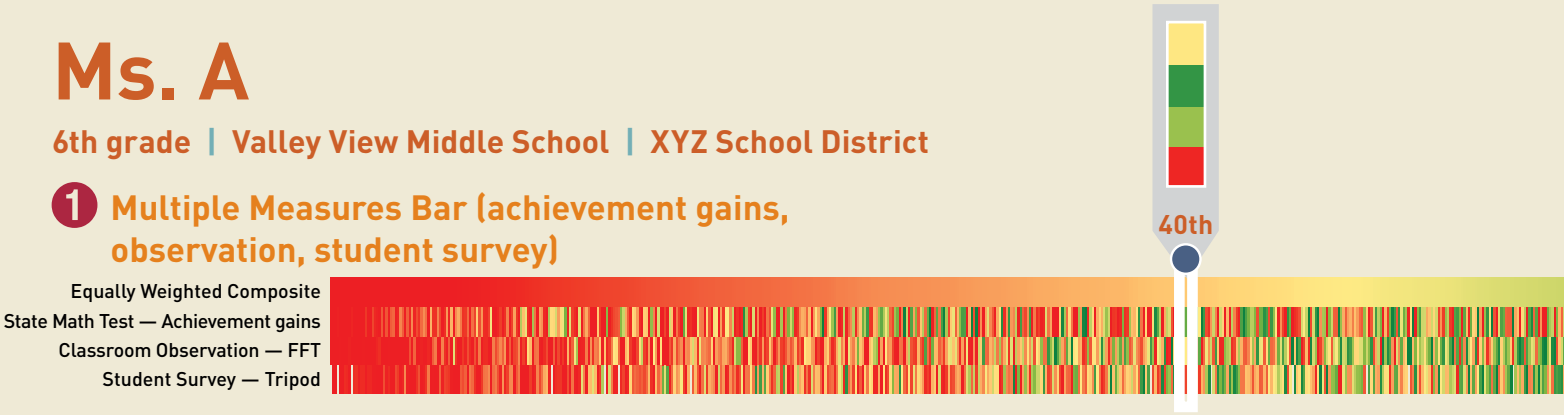
Diagnosing Practice with Multiple Measures

These pages use MET project data to illustrate how multiple measures can provide teachers with rich, contextualized information on their practice for use in professional development. Displayed are results for a MET project teacher (the name is fictional), her school, and district on classroom observations, student perception surveys, and student achievement gains. The teacher can see her overall results and where her results sit within the systemwide distribution for each measure and individual teaching competency.

Ms. A

6th grade | Valley View Middle School | XYZ School District

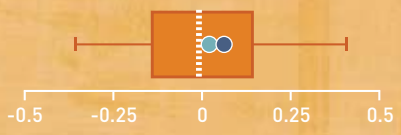
1 Multiple Measures Bar (achievement gains, observation, student survey)



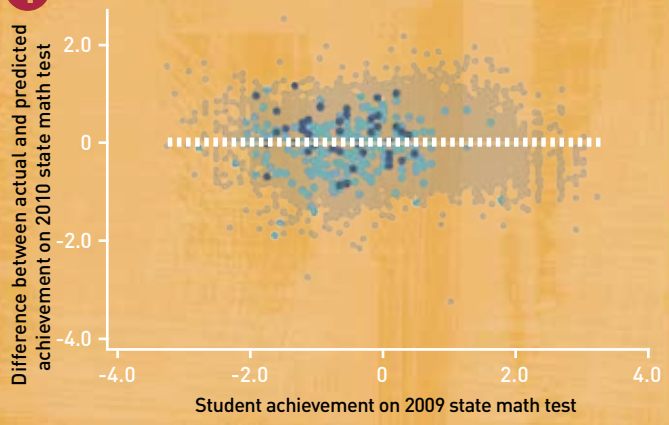
Achievement Gains

Middle School Math Scores

2



4

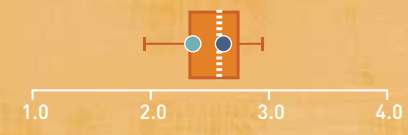


Students in Ms. A's ...

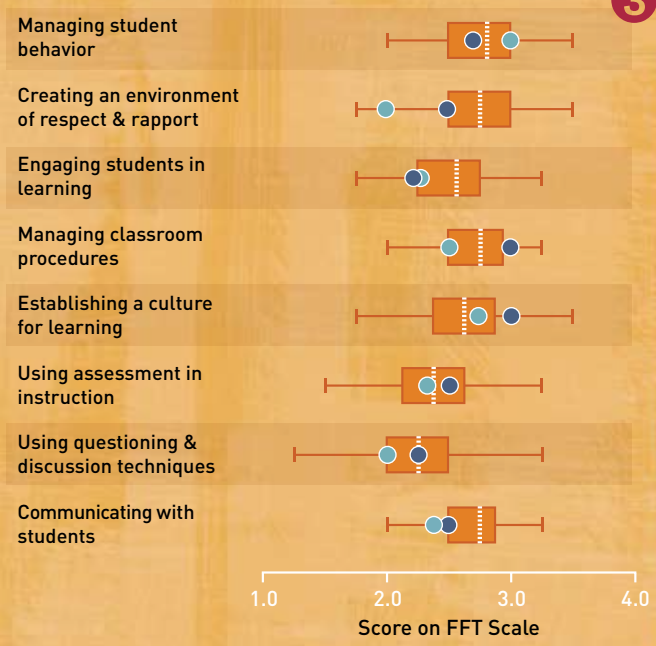
Classroom | School | District | Actual = Predicted Achievement

Classroom Observations

Score on Danielson Framework for Teaching (FFT)



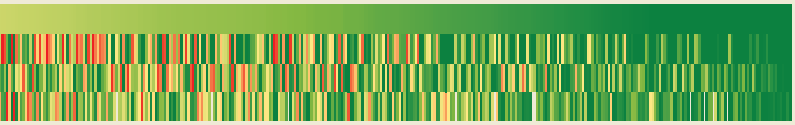
3

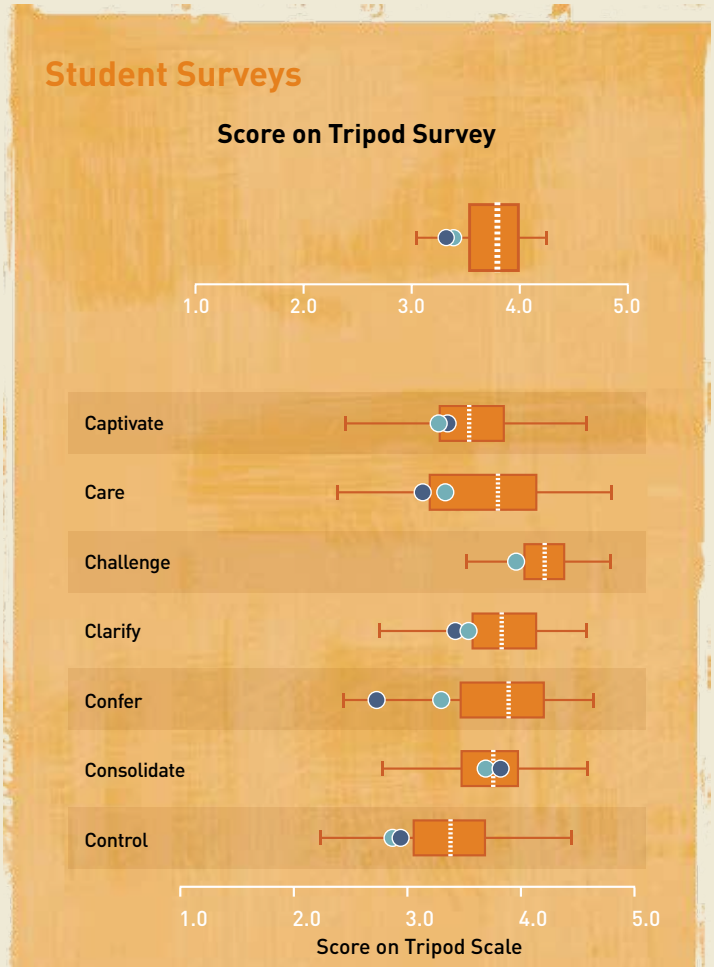


Composite score: **228 of 500**

Composite percentile: **40th**

Tier: **Satisfactory**





❶ Multiple Measures Bar

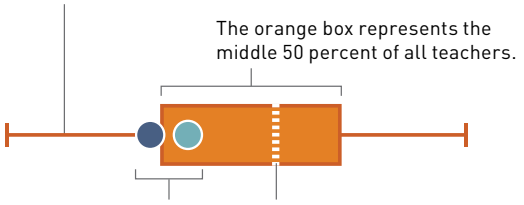
This bar contains a score for every teacher on each measure within the **Multiple Measures Composite** (MMC). The top row is the MMC and the **rows** below represent the achievement gains for the state math assessment and the average scores for the Framework for Teaching classroom observations and the Tripod student survey. Each **column** represents a single teacher. MMC scores determine placement on the bar from the lowest MMC score on the left to the highest MMC score on the right. Scores for the MMC and its individual measures are color-coded to performance standards for each measure, with red representing low performance, yellow representing average performance, and green representing high performance. Note that the colors generally match across the four measures near each end of the bar, indicating a high level of agreement among them. In other words, teachers at the very high end tend to do well on all of the measures, and the opposite is true for those at the very low end.

❷ & ❸ Box Plots

The box plots at level ❷ depict scores for each measure. The box plots at level ❸ depict scores for each component within the student survey and the teacher observation measures.

Legend

Lines extend from each side of the box, on the left to the 5th percentile and on the right to the 95th percentile. Scores beyond these lines are considered outliers.



❹ Achievement Gains Scatterplot

The scatterplot shows the gap between actual and predicted performance for all district 6th grade students on last year's state math assessment. The center (dashed) line represents actual performance equal to predicted performance. Predicted performance is the average performance for students with similar prior scores, after adjusting for English language learner and free and reduced-price lunch status. Points above the line represent higher-than-predicted performance for students with similar characteristics. Points below the line represent lower-than-predicted performance. Distance from the line represents the gap between predicted and actual performance. A teacher's value-added score is calculated by averaging each of his or her student's performance against predictions. Above predicted performance is credited as positive and below predicted performance is debited as negative.

continued from page 3

among teachers. Moreover, balanced weights avoid the risks posed by too narrow a focus on one measure. Overweighting any single measure (whether student achievement, observations, or surveys) invites manipulation and detracts attention and effort away from improvement on the other measures.

Ensure High-Quality Data

- **Monitor Validity.** Teachers who demonstrate skills and score high on a measure should experience more success in helping students learn than teachers with lower scores. If the skill, as measured, does not lead to better student outcomes, then it is not the best place for teachers to focus their limited time and attention. We have tested the validity of all measures in the MET project and found that students learn better in the classrooms of teachers with better observation scores, better student survey results, and prior success raising student test scores (adjusted for students' different starting points). One measure of content knowledge for teaching did not pass our validity test and was therefore omitted from our composite measure.

  The MET project invested considerable effort to randomly assign classrooms of students to teachers to determine if measures could identify effective teachers regardless of student assignment—and they did. School systems needn't go to such lengths, but they should compare teachers' performance on each measure with their student achievement gains. Validation is not a one-time exercise. If teachers begin to score higher on a measure, but higher scores are no longer associated with desired outcomes, then new measures are needed.

- **Ensure Reliability.** Low reliability indicates measurement error, and this undermines trust in the system. Measurement of teaching should reflect the quality of teachers' practice and not the idiosyncrasies of a particular lesson, observer, or group of students. The MET project has learned much about how to reliably measure teacher practice. Chief among these is the need to observe more than one lesson and include more than one observer for each teacher. In addition, we learned that short observations to supplement full-lesson observations can increase the reliability of observation ratings.

  School systems can use a variety of combinations of observers and lessons observed to improve reliability. For example, we found that school systems could achieve reliability above 0.65 when a principal observes one full lesson and peers or other administrators observe three partial lessons. The above scenario is more efficient, yet it achieves the same reliability as when a principal observes two full lessons and a peer or another administrator observes two additional full lessons. For student survey measures and tests, reliability is a function of the content of the questions, the consistency of the data collection process, and for survey questions, assurance of confidentiality.

- **Assure Accuracy.** Reliability without accuracy amounts to being consistently wrong. Because two observers agree, it does not mean they are correct. Accuracy of observations requires rigorous training on how to differentiate performance across all competencies within an observation instrument. It also requires assessment of observers' abilities to apply the instrument as intended before they are allowed to rate teachers' practice. Assuring accuracy of student test scores and survey responses means crediting them to the right teacher. Whenever the MET project collected student data from a classroom it verified with the teacher the names of the students in the class. School systems should do the same.

Invest in Improvement

- **Make Meaningful Distinctions.** Many traditional evaluation systems told almost all teachers they were satisfactory, and told very few, sometimes less than 1 percent, they were not. This does not reflect reality, but neither does a system that separates teachers into four equal-sized groups. MET project data suggest that teachers' effectiveness is unlikely to be distributed equally among several performance categories. Indeed, we found this to be far

from the case. MET project teachers' classroom observation scores were bunched at the center of the distribution, where 50 percent of the teachers scored within 0.4 points of each other (on a four-point scale) using Charlotte Danielson's Framework for Teaching. Teachers at the 25th and 75th percentiles scored less than one-quarter point different from the average. Only 7.5 percent of teachers scored below a two, and only 4.2 percent of teachers scored above a three. This would suggest a large middle category of effectiveness with two smaller ones at each end. Rather than trying to make fine distinctions among teachers in this vast middle, efforts would be better spent working to improve their practice.

- **Prioritize Support and Feedback.** While some teachers' low performance will require administrative action on behalf of students, it's a waste of effort to use measures of teaching only for high-stakes decisions. Multiple measures provide rich information to help teachers improve their practice. Although we didn't study the effectiveness of feedback, many of the teachers who participated in the MET project video study told us that seeing themselves teach was one of their most valuable professional development experiences.

- **Use Data for Decisions at All Levels.** The responsibility for improving teaching shouldn't rest with teachers alone. Measures of effective teaching enable school systems to better support teachers' improvement needs. Sound measures help school systems know where to target professional development and whether the supports work. A number of our partner districts—including the Denver and Hillsborough County (Fla.) Public Schools—have shifted professional development resources to areas of teaching that classroom observation measures indicate need improvement most. This led Hillsborough County to focus its professional development support on rigorous instructional techniques—for which teachers showed the most room for improvement—rather than classroom management skills that most teachers had clearly mastered.

The Next Phase of Work

States and districts have learned a great deal in the last few years about how to create better teacher development and evaluation systems. But there's still much to learn as these systems are implemented and improved over time and aligned to new expectations for students. One of the most exciting prospects is aligning teacher development and evaluation systems to the Common Core State Standards. As they move forward, states and districts should commit to measurement but hold lightly to the specific measures as the field continues to gain new knowledge. Understanding how teachers are performing is an important first step. But the real work lies ahead: understanding how to use that data to help all teachers improve their practice and the outcomes for America's young people.

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.

©2013 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