Testimony of Steve Cantrell
Chief Research Officer, Bill & Melinda Gates Foundation

Before the House Subcommittee on
Early Childhood, Elementary, and Secondary Education

February 28, 2013

Chairman Rokita, Ranking Member McCarthy, and committee members, I am Steve Cantrell, Chief Research Officer at the Bill & Melinda Gates Foundation and co-Director of the Measures of Effective Teaching project. Thank you for inviting me to testify about the MET project, a research study with great relevance for those who design and implement teacher evaluation and feedback systems.

Overview

The Measures of Effective Teaching project set out to test if it is possible to measure teaching effectiveness using multiple measures of a teacher's performance. The answer is yes. Drawing upon data collected from nearly 3,000 teachers from over 300 schools across six urban school districts, MET researchers demonstrated that effective teaching causes better learning. There are teachers whose students consistently outperform their peers and teachers whose students consistently underperform their peers. MET proved that these results are due to differences in teaching ability rather than differences in student characteristics, and that more and less effective teachers can be identified through a combination of classroom observations, student surveys, and evidence of student learning. These measures have the potential to provide teachers with much better feedback and more tailored supports to improve their practice and to help their students succeed.

Data Collection and Findings

The study looked at several dimensions of teaching. This is important because, as you know, teaching is complex and any single measure cannot fully reflect all important aspects of teaching. We measured four distinct aspects of teaching practice. We used two different student assessments to measure student learning. We used five different classroom observation protocols to assess the quality of classroom teaching (we are, of course, not recommending that districts adopt five different protocols). We tested teachers' ability to represent, identify, and increase students' conceptual understandings. Finally, we surveyed students themselves to assess how they experience the instructional environment.

This work was conducted by some of the nation's finest researchers and technical assistance providers using state-of-the-art methods and technology. The researchers used a value-added model (VAM) to calculate the differences between the actual and predicted performance of a teacher's students on both state tests in math and ELA in grades 4 through 9 and an additional more cognitively challenging

assessment in the same grades and subjects. Classroom lessons were observed using panoramic video cameras and scored by highly trained and certified raters. The test of teacher knowledge and the student perception of the instructional environment survey both built upon more than a decade of prior research.

Preliminary MET findings demonstrated that three measures—student assessments, classroom observations, and student surveys—helped predict whether teachers would raise the performance of future groups of students. Indeed, the combination of these measures does a far better job predicting which teachers will succeed in raising student performance than master's degrees and years of teaching experience.

In the study's second year, researchers took the unusual step to randomly assign classes of students to teachers. We did this to see if teachers previously identified as more effective based on these measures actually caused students to learn more. Random assignment allowed researchers to isolate teaching effectiveness from any unmeasured student characteristics. Furthermore, the researchers detected no bias in the teacher effectiveness estimates, as long as the estimates were adjusted to account for differences in measured students' characteristics, such as prior performance and demographics.

Final MET findings literally proved that effective teachers cause their students to learn more. Furthermore, the final findings showed that when combining measures into a single composite index, balanced weights are best. Composites that weigh state test results between 33% and 50% are more stable from year to year and better predict student performance on higher order assessments than composites that place more than 50% of the emphasis on state tests.

Nine Principles for Feedback and Evaluation Systems

It is now time for school systems to put into practice MET's research findings by building and implementing feedback and evaluation systems using multiple measures that teachers can trust. The MET project's final report, Feedback for Better Teaching, provides 9 principles to guide school systems as they develop feedback and evaluation systems. These 9 principles fall into three categories: Measure Effective Teaching, Ensure High Quality Data, and Invest in Improvement.

As school systems set out to measure effective teaching, there are three important considerations. First, the measures should emerge from and help establish expectations for what constitutes effective teaching practice. Second, since no single measure of effectiveness can capture the full complexity of teaching, states and districts should use multiple measures. Third, our research demonstrated that balance is best when considering how much emphasis to place upon any one measure within a set of multiple measures.

As school systems collect effectiveness data, there are three important considerations for establishing and maintaining trust in the data. First, the measures

should be valid predictors of increased student learning. A school system enters into a bargain with its teachers when it adopts a measure within an evaluation system. The bargain states that if teachers work hard to improve on this measure, then their students will be better learners. It is this bargain that animates the feedback promise of multiple measures. By annually validating each measure, the school system guarantees that effort toward improving practice will not be wasted. Second, the measurement process should be reliable. Teachers have been especially wary of classroom observation processes because they perceive the process as potentially subjective. MET project research discovered three ways to increase reliability of classroom observation: test and certify raters, have at least two raters observe each teacher, and observe at least two lessons. Third, when data are used for accountability, it is essential that the data match the right teachers with the right students. If the data are mismatched then one could easily draw the wrong conclusion about the effectiveness of a given teacher or school.

As school systems use effectiveness data, it is important to understand and communicate that improvement is the goal. Relatively few teachers in the MET sample exhibited uniformly poor or great practice across all measures. The data led us to conclude that most teachers are average, but for different reasons. Indeed, the majority of teachers scored very close to the mean on both the classroom observation instruments and on the survey of students' perceptions of the instructional environment. Yet, we know that average teaching is not good enough to help students achieve college and career success, so improvement is necessary. The realization that most teachers are in the middle means that school systems need to share the responsibility to improve teaching by providing targeted, high quality support.

As school systems begin this work, there are three important considerations for signaling an improvement-focused feedback and evaluation system. First, a system built for improvement will not exaggerate small differences, but will use performance categories to make meaningful distinctions among teachers. Teachers in adjacent categories should have demonstrably different impacts on student learning. Otherwise, there is no need for the additional category. Second, a system built for improvement will prioritize feedback and support in all communications with stakeholders. Third, though measures of effective teaching naturally focus on classrooms, the data from these measures should be used for decision-making at all levels of the school system. The measures will indicate areas where teachers need better support and this data should be used to determine which professional development to offer to which teachers and whether the professional development investments in place are making a difference to improve teaching practice. Furthermore, the measures will indicate the schools where teaching is getting better over time. This seems like a natural indicator of the quality of instructional leadership.

In closing, I want to reiterate one important point: Better evaluation and feedback systems are essential to improving teaching and learning. If done well, in ways that

teachers can trust, these systems will enable better teacher supports which, in turn, will lead to better student performance.

Thank you again for the opportunity to present.

Reports to accompany the written testimony

Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study
http://metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Feedback for Better Teaching: Nine Principles for Using Measures of Effective Teaching
http://metproject.org/downloads/MET_Feedback%20for%20Better%20Teaching_Principles%20Paper.pdf