Can Measures of Effective Teaching Identify Teachers Who Better Help Students Learn?[2]

By definition, teaching is effective when it enables student learning. But identifying effective teaching is complicated by the fact that teachers often have very different students. Students start the year with different achievement levels and different needs. Moreover, some teachers tend to get particular types of students year after year (that is, they tend to get higher-performing or lower-performing ones). This is why so-called value-added measures attempt to account for differences in the measurable characteristics of a teacher's students, such as prior test scores and poverty.

"Teachers previously identified as more effective caused students to learn more. Groups of teachers who had been identified as less effective caused students to learn less."

However, students differ in other ways—such as behavior and parental involvement—which we typically cannot account for in determining teaching effectiveness. If those "unaccounted for" differences also affect student learning, then what seems like effective teaching may actually reflect unmeasured characteristics of a teacher's students. The only way to know if measures of teaching truly identify effective teaching and not some unmeasured student characteristics is by randomly assigning teachers to students. So we did.

In 2009–10, we measured teachers' effectiveness using a combined measure, comprising teachers' classroom observation results, student perception survey responses, and student achievement gains adjusted for student characteristics, such as prior performance and demographics. The following year (2010–11), we randomly assigned different rosters of students to two or more MET project teachers who taught the same grade and subject in the same school. Principals created rosters and the RAND Corp assigned them randomly to teachers (see **Figure 1**). Our aim was to determine if the students who were randomly assigned to teachers who previously had been identified as more effective actually performed better at the end of the 2010–11 school year.[3]

They did. On average, the 2009–10 composite measure of effective teaching accurately predicted 2010–11 student performance. The research confirmed that, as a group, teachers previously identified as more effective caused students to learn more. Groups of teachers who had been identified as less effective

caused students to learn less. We can say they "caused" more (or less) student learning because when we randomly assigned teachers to students during the second year, we could be confident that any subsequent differences in achievement were being driven by the teachers, not by the unmeasured characteristics of their students. In addition, the magnitude of the gains they caused was consistent with our expectations.

Figure 2 illustrates just how well the measures of effective teaching predicted student achievement following random assignment. The diagonal line represents perfect prediction. Dots above the diagonal line indicate groups of teachers whose student outcomes following random assignment were better than predicted. Dots below the line indicate groups of teachers whose student outcomes following random assignment were worse than predicted. Each dot

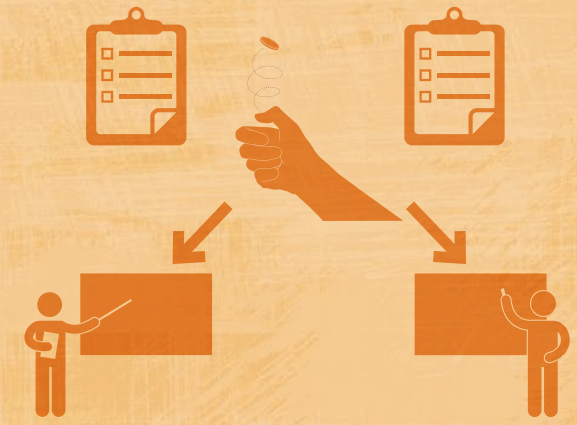
Figure 1

Putting Measures of Effective Teaching to the Test with Random Assignment

1. Principals created rosters for each class

2. The rosters were assigned randomly within each grade and subject

3. We predicted student outcomes based on teachers' previous results, observations, and student surveys.



4. We compared those predictions to actual differences.

Do measures of teaching really identify teachers who help students learn more, or do seemingly more effective teachers just get better students? To find out, the MET project orchestrated a large-scale experiment with MET project teachers to see if teachers identified as more effective than their peers would have greater student achievement gains even with students who were assigned randomly.

To do so, the MET project first estimated teachers' effectiveness using multiple measures from the 2009–10 school year. As is common in schools, some teachers had been assigned students with stronger prior achievement than others. In assessing each teacher's practice that year, the project controlled for students' prior achievement and demographic characteristics. But there may have been other differences among students as well. So for the following school year (2010–11), principals created rosters of students for each class in the study, and then researchers randomly assigned each roster to a participating teacher from among those who could teach the class.

At the end of the 2010–11 school year, MET project analysts checked to see if students taught by teachers identified as more effective than their colleagues actually had greater achievement gains than students taught by teachers identified as less effective. They also checked to see how well actual student achievement gains for teachers matched predicted gains.

represents 5 percent of the teachers in the analysis, sorted based on their predicted impact on student achievement.[4]

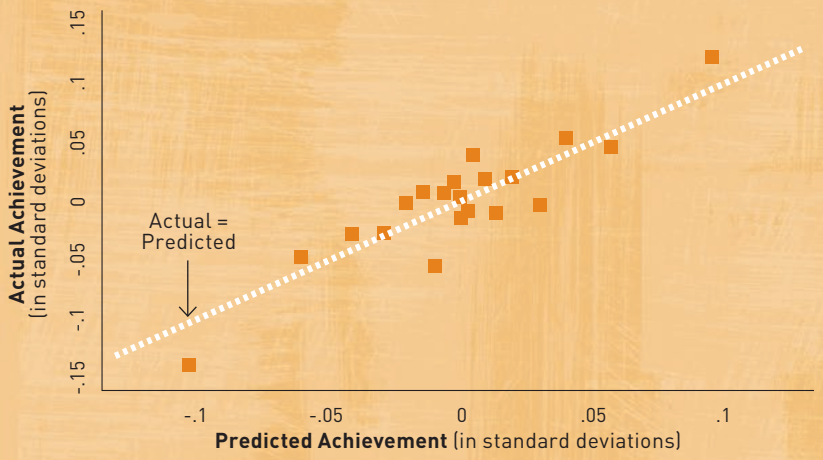
As seen in **Figure 2**, in both math and English language arts (ELA), the groups of teachers with greater predicted impacts on student achievement generally had greater actual impacts on student achievement following random assignment. Further, the actual impacts are approximately in line with the predicted impacts.[5] We also found that teachers who we identified as being effective in promoting achievement on the state tests also generated larger gains on the supplemental tests administered in spring 2011.

Based on our analysis, we can unambiguously say that school systems should account for the prior test scores of students. When we removed this control, we wound up predicting much larger differences in achievement than actually occurred, indicating that student assignment biased the results. However, our analysis could not shed as much light on the need to control for demographics or "peer effects"—that is, the average prior achievement and demographics of each student's classmates. Although we included those

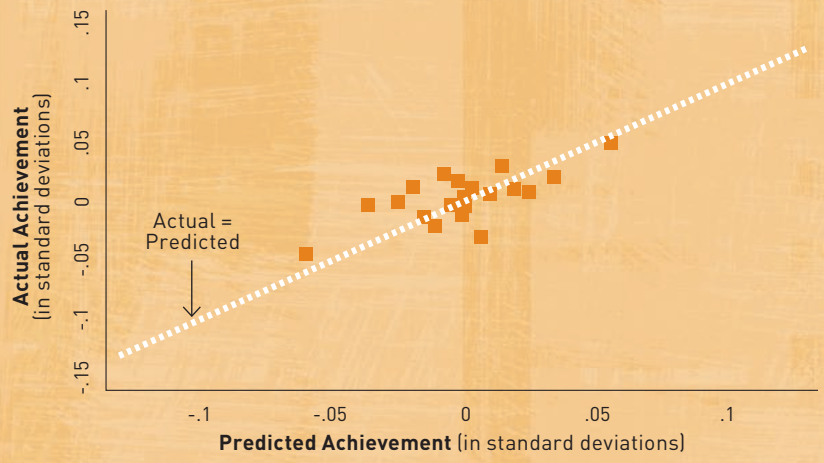
Figure 2

Effectiveness Measures Identify Teachers Who Help Students Learn More

Actual and Predicted Achievement of Randomized Classrooms (Math)



Actual and Predicted Achievement of Randomized Classrooms (English Language Arts)



These charts compare the actual 2010–11 school year achievement gains for randomly assigned classrooms with the results that were predicted based on the earlier measures of teaching effectiveness. Each dot represents the combination of actual and estimated student performance for 5 percent of the teachers in the study, grouped by the teachers' estimated effectiveness. The dashed line shows where the dots would be if the actual and predicted gains matched perfectly.

On average, students of teachers with higher teacher effectiveness estimates outperformed students of teachers with lower teacher effectiveness estimates. Moreover, the magnitude of students' actual gains largely corresponded with gains predicted by their effectiveness measured the previous year. Both the actual and predicted achievement are reported relative to the mean in the randomization block. That is, a zero on either axis implies that the value was no different from the mean for the small group of teachers in a grade, subject, and school within which class lists were randomized.

Impacts are reported in student-level standard deviations. A .25 standard deviation difference is roughly equivalent to a year of schooling. The predicted impacts are adjusted downward to account for incomplete compliance with randomization.

> "We can unambiguously say that school systems should adjust their achievement gain measures to account for the prior test scores of students. When we removed this control, we wound up predicting much larger differences in achievement than actually occurred."

controls, we cannot determine from our evidence whether school systems should include them. Our results were ambiguous on that score.

To avoid over-interpretation of these results, we hasten to add two caveats: First, a prediction can be correct on average but still be subject to measurement error. Our predictions of students' achievement following random assignment were correct on average, but within every group there were some teachers whose students performed better than predicted and some whose students performed worse. Second, we could not, as a practical matter, randomly assign students or teachers to a different school site. As a result, our study does not allow us to investigate bias in teacher effectiveness measures arising from student sorting between different schools.[6]

Nonetheless, our analysis should give heart to those who have invested considerable effort to develop practices and policies to measure and support effective teaching. Through this large-scale study involving random assignment of teachers to students, we are confident that we can identify groups of teachers who are comparatively more effective than their peers in helping students learn. Great teaching does make a difference.