

Can student surveys measure teaching quality?

The results are in: Surveys of K-12 students covering a range of classroom characteristics linked to teacher quality can successfully predict student achievement.

By Ronald F. Ferguson

Consider this: Once there was a classroom filled with busy students getting their work done. The principal often marveled at how well-behaved they were when she peered through the small window in the classroom door.

But, what she could not see was that most lessons were uninteresting, assignments emphasized memorization more than understanding, and many days the teacher seemed indifferent to students' feelings and opinions. It was not a happy place. There was no love of learning. The principal and the teacher believed that his teaching was great because his end-of-year test scores were almost always above average. Neither understood that anything was wrong. Neither asked the students.

In higher education, colleges and universities rely on students to evaluate teaching. Student responses to teacher evaluations help administrators make pay and promotion decisions, and enlighten fellow students about courses to take or avoid. But, can students in primary and secondary schools help evaluate teaching?

They spend hundreds more hours in each classroom than any observer ever will. Nonetheless, until now, school improvement efforts have seldom sought systematic student feedback at the classroom level (as opposed to the whole school level) in primary and secondary schools. One impediment has been the doubt that students can provide valid and reliable responses about the quality of the teaching that they experience.

RONALD F. FERGUSON (ronald_ferguson@harvard.edu) is a senior lecturer in education and public policy at the Harvard Graduate School of Education and the Harvard Kennedy School, Cambridge, Mass.

The Measures of Effective Teaching (MET) Project of the Bill & Melinda Gates Foundation tests whether this doubt is well-founded. For 3,000 teachers, the MET project examines multiple measures of teaching, including value-added test score gains, observational protocols, professional knowledge tests, and student perception surveys. To measure student perceptions, MET selected the Tripod student survey that I developed in my consulting work with public schools. Tripod refers to content knowledge, pedagogic skill, and relationships.

From this work, we have learned that well-crafted student surveys can play an important role in suggesting directions for professional development and also in evaluating teacher effectiveness. However, there are some very important cautions to keep in mind. First, any method of measuring a classroom is prone to measurement error. Second, and perhaps most importantly, if a single deployment of a survey or observational protocol can have major consequences, then many teachers may temporarily alter their behaviors during the measurement period to try to influence the outcome. This has the potential to make the measurements invalid. For both reasons, a wise rule to follow is that no one student survey or classroom observational protocol — and no single deployment of any one measure — should be used by itself to have a large effect on any teacher evaluation decision.

I join many others in believing that the way forward for the field of teacher evaluation *and support* should be “multiple measures, multiple times, over multiple years.” And student surveys should be one of those measures.

Background

The first few generations of the Tripod survey (from 2001-05) were refined in consultation with K-12 teachers and administrators in Shaker Heights, Ohio, and 15 member districts of the Minority Student Achievement Network.

Since then, multiple surveys geared for different age groups have continued evolving and the work has spread around the nation and even abroad. From 2001 through 2012, almost a million elementary, middle, and high school students in the U.S. completed Tripod surveys. Items have been added and replaced each year to gradually develop a stable set of valid and reliable indices. By responding to these surveys, students have provided feedback to their schools and teachers not only about their schools as whole institutions, but also in the specific classrooms where they took the surveys. Teachers receive personalized summaries of the responses from their own classrooms. By prior agreement, reports for administrators sometimes hide teachers' identi-

ties. In circumstances where surveys are part of an evolved teacher evaluation system, supervisors can sometimes see results at the teacher level. In either case, the results show patterns of instructional disparity between classrooms and associated variation in student engagement, which can inform evaluation, and decisions about professional development and school improvement.

Measures of teaching quality

The primary measures of instructional quality in the Tripod surveys are gathered under seven headings called the Tripod 7C's: Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate. They are grounded in the work of many education researchers over several decades. They capture much of what researchers have suggested is important in determining how well teachers teach and how much students learn. Each of the C's is measured using multiple survey items. The following are brief descriptions of each concept.

Well-crafted student surveys can play an important role in suggesting directions for professional development and also in evaluating teacher effectiveness.

The Tripod 7C's

Care pertains to teacher behaviors that help students feel emotionally safe and to rely on the teacher to be an ally in the classroom. Caring reduces anxiety and provides a sense of positive affiliation and belonging. Caring goes beyond “niceness”; caring teachers work hard and go out of their way to help. They signal to their students, “I want you to be happy and successful, and I will work hard to serve your best interest; your success is an important source of my personal satisfaction.” An example of a Tripod survey item measuring Care is: “My teacher really tries to understand how students feel about things.”

Control pertains to classroom management. Teachers need skills to manage student propensities toward off-task or disruptive behaviors in order to foster conditions that allow for effective communication and focus. Control helps to maintain order and supplements caring by making the classroom calm and emotionally safe from such things as negative peer pressure. An example of a Tripod survey item measuring Control is: “Our class stays busy and doesn't waste time.”

Clarify concerns teacher behaviors that promote understanding. Interactions that clear up confusion and help students persevere are especially important.

Each student comes with particular gaps in understanding and with both correct and incorrect interpretations of the world around them. To be most effective, teachers should be able to diagnose students' skills and knowledge, and they need multiple ways of explaining ideas that are likely to be difficult for students to grasp. Teachers also must judge how much information students can absorb at any one time, and they should differentiate instruction according to individual maturity and interest. An example of a Tripod survey item measuring Clarify is: "My teacher has several good ways to explain each topic that we cover in this class."

Challenge concerns both effort and rigor — pressing students to work hard and to think hard. Challenging teachers tend to monitor student effort and to confront students if their effort is unsatisfactory. Students who do not devote enough time to their work or who give up too easily in the face of difficulty are pushed to do more. Similarly, students who do not think deeply or who resist reasoning their way through challenging questions are both supported and pushed. The teacher may ask a series of follow-up questions intended to elicit deeper, more thorough reasoning. An example of a Tripod survey item measuring Challenge for effort is: "In this class, my teacher accepts nothing less than our full effort." An item measuring Challenge for rigorous thinking is: "My teacher wants us to use our thinking skills, not just memorize things."

Captivate concerns teacher behaviors that make instruction stimulating instead of boring. Captivating teachers make the material interesting, often by demonstrating its relevance to things about which students already care. Brain research establishes clearly that stimulating learning experiences and relevant material make lessons easier to remember than when the experience is boring and the material seems irrelevant. Examples of survey items that measure stimulation and relevance are: "My teacher makes lessons interesting," and "I often feel like this class has nothing to do with real life outside school."

Confer concerns seeking students' points of view by asking them questions and inviting them to express themselves. When students expect that the teacher might call on them to speak in class, they have an incentive to stay alert and engaged. In addition, believing that the teacher values each point of view provides positive reinforcement for the effort required to formulate a perspective in the first place. Further, if students are asked to respond not only to the teacher but also to one another, a learning community may develop in the classroom with all the attendant social reinforcements. An example of an item for Confer is: "My teacher gives us time to explain our ideas."

Consolidate concerns how teachers check for understanding and help students organize material for more effective encoding in memory and for more efficient reasoning. These practices include reviewing and summarizing material at the end of classes and connecting ideas to material covered in previous lessons. Teachers who excel at consolidation talk about the relationships between ideas and help students see patterns. There is a large body of evidence supporting the hypothesis that these types of instructional activities enhance retention by building multiple mental pathways for retrieving knowledge and for combining disparate bits of knowledge in effective reasoning. An example of a survey item measuring Consolidation is: "My teacher takes the time to summarize what we learn each day."

MET findings suggest that the highest-achieving classrooms are respectful and orderly environments, with students who stay busy and learn to correct their mistakes from a teacher who explains difficult things clearly.

The Tripod survey combines between four and eight survey items associated with each particular C into a composite index for each concept. The result is seven composite indices, one for each of the C's.

There is a useful distinction in the teaching quality literature between "support" and "press." Support refers to ways of serving and communicating with students that the students are likely to perceive as helpful and welcome. Conversely, press imposes some degree of stress. It refers to communication that students may experience as pressure to move outside of their comfort zones. We distinguish the Five Support C's — i.e., Care, Clarify, Captivate, Confer, and Consolidate — from the two C's that entail a higher degree of press, i.e., Control and Challenge. Challenge and Control strongly connote teacher demands. However, our findings indicate that Challenge may also entail aspects of support, making it harder to classify in just one category.

How much do classrooms differ?

The Tripod 7C's distinguish classrooms quite markedly. To provide a sense of just how much, we examine variation in the percentage of favorable responses to 7C's items.

Even for early elementary school classrooms, there is a great deal of between-class variation in how favorably students rate their teachers. Most of the variation occurs within schools, rather than between them. Most schools have both high- and low-rated classrooms. The question is whether the ratings are

systematic and meaningful enough to inform teacher evaluation and school improvement.

Student responses to the Tripod 7C's are indeed valid and reliable predictors of learning in mathematics and English language arts, according to MET reports. The January 2012 report in particular indicated that Tripod student surveys were substantially more reliable predictors of value-added achievement gains than classroom observations. Earlier, the December 2010 report had established that differences between teachers in how highly their students had rated them predicted important differences in how much students learned.

MET's December 2010 report ranks teachers based on their student survey responses, then compares how much students learn in classes taught by teachers that students rate high compared to those that they rate low. One version of the analysis correlates survey responses with learning gains in other sections taught by the teacher during the same school year. Another examines gains in classrooms taught in the prior year. In each analysis, students of math teachers with Tripod survey rankings in the top quarter learned the equivalent of 4 to 5 months more per year, on average, than students of teachers with survey rankings in the bottom quarter. The differences for English language arts were also significant, with gains about half as large as for math.

Educators in the audiences that I speak to usually expect that Care will be the strongest predictor of value-added achievement gains. However, that is not borne out by the MET project. Table 1 shows the six items that most strongly predict gains in the MET data.

MET findings suggest that the highest-achieving classrooms are respectful and orderly environments, with students who stay busy and learn to correct their mistakes from a teacher who explains difficult things clearly. Control is the strongest predictor. However, the differences between these correlations and those for other items are not large. All of the Tripod 7C's items and associated indices are correlated in the MET study with value-added achievement gains. Educators should keep all of them in mind as they seek ways to improve teaching and learning.

That student perceptions of teaching quality have now been shown to predict test scores is important, but test scores are not the only outcomes that we value. As parents, teachers, and concerned citizens, we also value lofty goals, on-task behaviors, optimistic beliefs, and happy feelings.

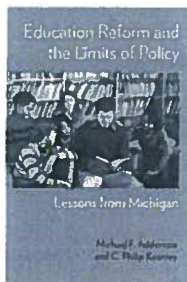
In our work with the Tripod survey in other projects, we have learned that students are happier in classrooms where they feel safe and intellectually stimulated — classrooms where teachers relate well to students, make lessons interesting, and maintain

TABLE 1.
Tripod 7 C's responses

Rank	Survey statement	Category
1	Students in this class treat the teacher with respect.	Control
2	My classmates behave the way my teacher wants them to.	Control
3	Our class stays busy and doesn't waste time.	Control
4	In this class, we learn a lot every day.	Challenge
5	In this class, we learn to correct our mistakes.	Challenge
6	My teacher explains difficult things clearly.	Clarify

Source: Bill & Melinda Gates Foundation. (2010, December). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project.*

Books that offer answers



Education Reform and the Limits of Policy Lessons from Michigan

Michael F. Addonizio
C. Philip Kearney

By examining a major set of education policy reforms undertaken in Michigan and across the country over the past 20-plus years, the authors are able to reveal the varying success of innovations such as finance reform, state assessment of student performance, school accountability measures, charter schools, and schools of choice.
\$40 cloth 978-0-88099389-0 | \$18 pbk. 978-0-88099-387-6



The Power of a Promise Education and Economic Renewal in Kalamazoo

Michelle Miller-Adams

"The Kalamazoo Promise is a fascinating experiment with enormous implications for American education. Will the promise of free tuition inspire students to work harder in school? Will it draw more middle class families back into the city's public school system, enriching the learning environment for all? Miller-Adams provides a superb, thoroughly researched, and highly readable account of this closely watched, groundbreaking venture."

—Richard D. Kahlenberg

\$40 cloth 978-0-88099-340-1 | \$18 pbk 978-0-88099-339-5.

W.E. UPJOHN INSTITUTE

FOR EMPLOYMENT RESEARCH

<http://www.upjohn.org>

ph. 888-227-8569 • fax 269-343-7310

order. Elementary school students try hardest to understand their lessons when teachers are not only caring, but when they also have the skills to provide clear explanations and challenge students to think and persist in the face of difficulty. Even for the high school students, two of the strongest three predictors of effort are highly favorable responses to Clarify and Challenge.

Students hide effort more and hold back more in classrooms where teachers confer more actively with their students. Perhaps students have more opportunities to avoid participating (by hiding effort and holding back) in classes where there is lots of conferring and opportunities to participate are more abundant. The fact that students are more aware of hiding effort or holding back in such classes may not necessarily be a bad thing. It's just human nature. Similarly, both upper and lower elementary school students agree that their teachers accuse them of not paying attention in classrooms where teachers work most actively with students to help them consolidate their knowledge. An example of an elementary school item for Consolidate is, "My teacher takes time to help us remember what we learn." If teachers who take time to help students remember are also more prone to notice when they are not paying attention, that is a good thing.

It appears that each of the Tripod 7C's pertains to



"You must be the new hire. Welcome aboard. Here's our prepackaged curriculum. If time permits, encourage critical thinking."

a distinct and potentially important aspect of teaching that affects what students experience and how actively they engage. In addition, there are multiple dimensions of engagement that we care about. We don't just want high test scores. We also want attentiveness and good behavior, happiness, effort, and efficacy. I think most parents would be willing to sacrifice a few test score points if it meant gaining more of these four. Fortunately, the same teaching behaviors that predict better behavior, greater happiness, more effort, and stronger efficacy also predict greater value-added achievement gains.

Educators and researchers worry that student survey responses to teacher quality questions will reflect mainly family backgrounds and personalities instead of teaching quality. So, we take pains to find ways to reduce the possibility that such extraneous factors affect our findings. This is why the MET project used the student survey responses from one set of classrooms to predict value-added gains in other classrooms taught by the same teachers. With this research design, the teacher is presumably the only factor affecting both the survey responses and the test score gains, so that any correlation between the two is presumably attributable to the teaching.

Conclusion

Researchers over many decades have suggested that students will engage more deeply and master their lessons more thoroughly when their teachers care about them, manage the classroom well, clarify complex ideas, challenge them, make lessons interesting, confer with them and consolidate lessons to make learning coherent. Using the Tripod 7C's survey, MET and other projects have given students in thousands of classrooms the opportunity to demonstrate that they can predict their own value-added achievement gains and distinguish their classrooms on key dimensions.

Doubts about whether student responses can be reliable, valid, and stable over time at the classroom level are being put to rest. We are learning that well constructed classroom-level student surveys are a low burden and high-potential mechanism for incorporating students' voices in massive numbers into our efforts to improve teaching and learning. No one survey instrument or observational protocol should have high stakes for teachers if used alone or for only a single deployment. But, students know good instruction when they experience it as well as when they do not. I anticipate a growing consensus that student perception surveys should be among the multiple measures used multiple times over multiple years not only to measure teaching effectiveness, but also to inform a broader set of actions that we can take to improve schools. ■