MET project

POLICY AND PRACTICE BRIEF

Asking Students about Teaching

Student Perception Surveys and Their Implementation

BILL & MELINDA GATES foundation

ABOUT THIS REPORT: Based on the experience of the Measures of Effective Teaching (MET) project, its partners, and other districts, states, and organizations, this report is meant for practitioners and policymakers who want to understand student surveys as potential tools for teacher evaluation and feedback, as well as the challenges and considerations posed by their implementation. MET project findings on the reliability and validity of student surveys are explored in depth in a separate research report, *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*.

ACKNOWLEDGMENTS: This document is based on the collective knowledge of many individuals who shared their experiences with student surveys for the benefit of their peers in other states, districts, and organizations. Thanks to:

- Ryan Balch, Vanderbilt University
- Kelly Burling and colleagues, Pearson
- Carrie Douglass, Rogers Family Foundation
- Ronald Ferguson, Harvard University and the Tripod Project for School Improvement
- Tiffany Francis, Sam Franklin, and Mary Wolfson, Pittsburgh Public Schools
- James Gallagher, Aspire Public Schools
- Alejandro Ganimian, Harvard University
- Hilary Gustave, Denver Public Schools
- Monica Jordan and Rorie Harris, Memphis City Schools
- Kevin Keelen, Green Dot Public Schools
- Jennifer Preston, North Carolina Department of Public Instruction
- Rob Ramsdell and Lorraine McAteer, Cambridge Education
- Valerie Threlfall and colleagues, Center for Effective Philanthropy
- Ila Deshmukh Towery and colleagues, TNTP
- Deborah Young, Quaglia Institute for Student Aspirations
- Sara Brown Wessling, Johnston High School in Johnston, Iowa

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. For more information, visit www.metproject.org.

September 2012

Contents

Why Ask Students about Teaching?

No one has a bigger stake in teaching effectiveness than students. Nor are there any better experts on how teaching is experienced by its intended beneficiaries. But only recently have many policymakers and practitioners come to recognize that—when asked the right questions, in the right ways—students can be an important source of information on the quality of teaching and the learning environment in individual classrooms.

A New Focus on the Classroom

As states and districts reinvent how teachers are evaluated and provided feedback, an increasing number are including student perception surveys as part of the mix. In contrast to the kinds of school climate surveys that have been commonly used in many systems, these student perception surveys ask students about specific teachers and specific classrooms. District-wide administration of such surveys has begun in Denver, Memphis, and Pittsburgh. Statewide pilots have taken place in Georgia and North Carolina. And student surveys are now used for feedback and evaluation by the **New Teacher Center**, **TNTP** (formerly The New Teacher Project), and such charter management organizations as **Aspire Public Schools** and **Green Dot Public Schools**.

There are many reasons for this new interest. Teaching is a complex interaction among students, teachers, and content that no one tool can measure. The search for different-but-aligned instruments has led many to use student surveys as a complement to such other tools as classroom observations and measures of student achievement gains.

Analysis by the Measures of Effective Teaching (MET) project finds that teachers' student survey results are predictive of student achievement gains. Students know an effective classroom when they experience one. That survey results predict student learning also suggests surveys may provide outcome-related results in grades and subjects for which no standardized assessments of student learning are available.

Further, the MET project finds student surveys produce more consistent results than classroom observations or achievement gain measures (see the MET project's *Gathering Feedback* policy and practitioner brief). Even a high-quality observation system entails at most a handful of classroom visits, while student surveys aggregate the impressions of many individuals who've spent many hours with a teacher.

Student surveys also can provide feedback for improvement. Teachers want to know if their students feel sufficiently challenged, engaged, and comfortable asking them for help. Whereas annual measures of student achievement gains provide little information for improvement (and generally too late to do much about it), student surveys can be administered early enough in the year to tell teachers where they need to focus so that their current students may benefit. As feedback tools, surveys can be powerful complements to other instruments. Surveys might suggest students' misunderstandings aren't being addressed; done well, observations can diagnose a teacher's specific attempts at clarification.

Imperatives for Implementation

Benefits aside, student perception surveys present their own set of challenges and considerations to states and districts. Some of these relate to the survey instrument itself. Not every survey will produce meaningful information on teaching. Not to be confused with popularity contests, well-designed student perception surveys capture important aspects of instruction and the classroom environment. Rather than pose, "Do you like your teacher?" typical items might ask the extent to which students agree that, "I like the way the teacher treats me when I need help" and "If you don't understand something, my teacher explains it another way." Survey development is a sophisticated, data-driven process.

But even a good instrument, implemented poorly, will produce bad information. Attending to issues such as student confidentiality, sampling, and accuracy of reporting takes on greater urgency as systems look toward including student surveys in their evaluation systems. The care with which systems must administer surveys in such contexts is akin to that required in the formal administration of standardized student assessments. Smooth administration and data integrity depend on piloting, clear protocols, trained coordinators, and quality-control checks.

The following pages draw on the experience of the MET project, its partners, and other leading school systems and organizations to clarify four overriding requirements of any system considering student surveys as part of formal feedback and evaluation for teachers:

1. **Measure what matters.** Good surveys focus on what teachers do and on the learning environment they create. Surveys should reflect the theory of instruction that defines expectations for teachers in a system. Teachers with better survey results should also have better outcomes on measures of student learning.

2. **Ensure accuracy.** Student responses should be honest and based on clear understanding of the survey items. Student confidentiality is a must. Accuracy also means that the right responses are attributed to the right teacher.

3. **Ensure reliability.** Teachers should have confidence that surveys can produce reasonably consistent results and that those results reflect what they generally do in their classrooms—not the idiosyncrasies of a particular group of students. Reliability requires adequate sampling and an adequate number of items—but without overtaxing students.

4. **Support improvement.** Measurement for measurement's sake is wasted effort. Teachers should receive their results in a timely manner, understand what they mean, and have access to professional development resources that will help them target improvement in areas of need. Student surveys are as much about evaluating systems of support for teachers as they are about diagnosing the needs within particular classrooms.

Addressing these requirements amid finite resources and competing concerns necessarily involves trade-offs. Although the study of student surveys has intensified, research to date cannot offer hard and fast rules for striking the right balance on many issues. Highlighted throughout this report are some of the varied approaches that systems are taking. But implementation in the field is a work in progress. Individual districts and states are piloting before deciding how to deploy surveys formally and at scale. This document concludes by emphasizing the importance of stakeholder engagement for achieving positive results.

It will be important to monitor how student surveys behave as they move into the realm of formal evaluation. The Tripod survey studied by the MET project, developed by Harvard researcher Ronald Ferguson, has been refined over 11 years of administration in many thousands of classrooms as a research and professional development tool. But only recently have systems begun to consider its use as part of formal teacher evaluation.

Any instrument, when stakes are attached, could distort behavior in unwanted ways, or produce a less accurate picture of typical practice. One can imagine a teacher who, consciously or not, acts more lenient if student surveys are factored into evaluation, even though well-designed surveys stress a balance of challenge and support. This is further reason for multiple measures. It lets teachers stay focused on effective teaching and not on any one result.

Benefits to Student Perception Surveys

1. **Feedback.** Results point to strengths and areas for improvement.
2. **"Face validity."** Items reflect what teachers value.
3. **"Predictive validity."** Results predict student outcomes.
4. **Reliability.** Results demonstrate relative consistency.
5. **Low cost.** Expense of administration is minimal.

Measure What Matters

If the objective of measurement is to support improved student outcomes, then the measures used should relate to those outcomes—and to the means teachers use to help students achieve them. It's a problem if a measure has no relationship to how teachers in a system are expected to teach, or to the outcomes for which they're held accountable. Making sure those relationships exist should be part of adopting a student perception survey.

Indicators of Teaching

To measure something broad like teaching requires breaking the activity down into discrete components that together represent a theory about the instruction that supports student learning. Developers of student perception surveys often refer to these components as "constructs"—different aspects of teaching and the learning environment, the quality of which are indicated by responses to multiple related questions about students' observations and feelings.

Consider the Tripod survey employed in the MET project study. Tripod is designed to measure teaching, student engagement, school norms, and student demographics. To measure teaching, the survey groups items under seven constructs, called the "7 Cs": Care, Control, Challenge, Clarify, Confer, Captivate, and Consolidate. For each, the survey poses a series of statements, asking students' level of agreement on a five-point scale. Here are two items under "Clarify":

- "My teacher has several good ways to explain each topic that we cover in class."
- "My teacher knows when the class understands and when we do not."

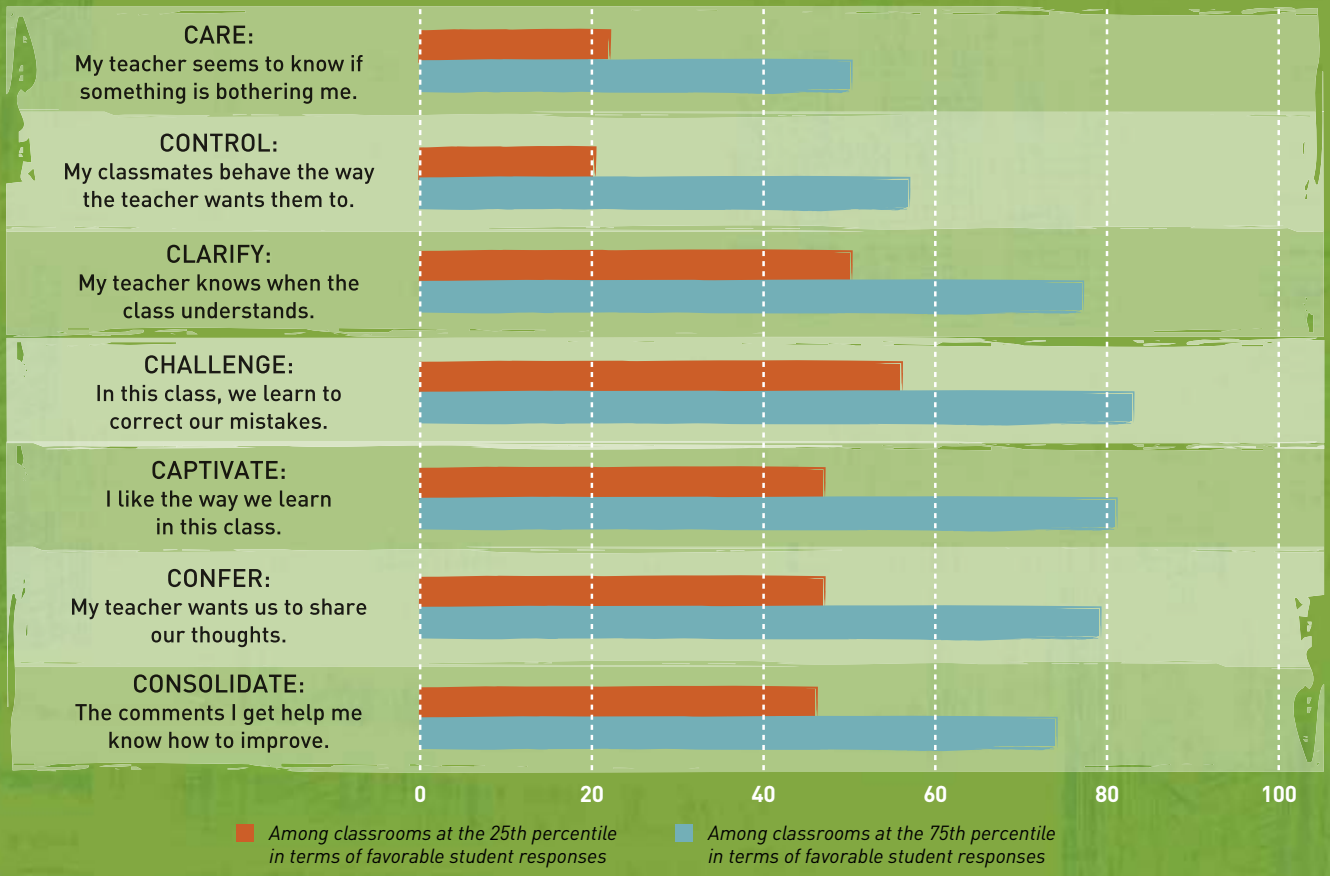
An important quality of any measure of teaching is its ability to differentiate among classrooms. If all teaching looks the same, the instrument is not providing much information for feedback. As shown in **Figure 1,** students who completed the Tripod survey as part of the MET project perceived clear differences among teachers. The chart shows the percentage of students agreeing with one statement from each construct among classrooms at the 75th and 25th percentiles in terms of how favorably their students responded.

Note the results for the item: "The comments I get help me know how to improve." Among classrooms at the 75th percentile in terms of favorable responses to that statement, more than three-quarters of students agreed. Among classrooms at the 25th percentile, well under half of students agreed. These results don't show how survey results relate to student achievement—that is addressed further on—but the contrasts suggest that students are capable of discerning

Figure 1

Students Perceive Clear Differences among Teachers

Percentage of students agreeing with each statement



An important aspect of any measure of teaching is its ability to distinguish among classrooms. Measures that fail this test provide little information for feedback. The above chart shows the percentage of students agreeing with one statement from each of the Tripod student perception survey's constructs among classrooms ranked at the 75th and 25th percentiles in terms of how favorably their students responded. Clearly, the students were able to discern significant differences among classrooms in terms of the qualities measured by the survey.

significant differences among classrooms.

Tripod is one of the oldest and most widely used off-the-shelf survey instruments, and the only one studied by the MET project. Some school systems have drawn from Tripod items included in the MET project as they adopt student surveys. But other tools exist. As shown in the table on pages 7 and 8—highlighting four currently available off-the-shelf surveys that ask about specific classrooms—instruments share many similar structures and constructs. Differences in emphasis often relate to a survey's particular origins. Many specifics on the number of items and their wording are moving targets as developers adjust their instruments based on ongoing research and system needs.

Student Perception Surveys with Classroom-Level Items

Tool and Background	Number of Items	Constructs	Response Options	Sample Items
Tripod (www.tripodproject.org) ▪ Developed by Harvard researcher Ronald Ferguson; distributed and administered by Cambridge Education ▪ Online and paper administration ▪ Grew out of study of student engagement and related teaching practices ▪ Versions for three grade bands: K–2; 3–5; 6–12 ▪ MET project study found Tripod predictive of achievement gains and able to produce consistent results for teachers (see MET project's *Learning about Teaching*)	▪ Core are approx. 36 items in the "Tripod 7 Cs" at the secondary level; fewer at earlier grades ▪ Additional items ask about student's engagement, background, and academic beliefs ▪ Full versions includes 80+ items; shorter forms available	**7 Cs** 1. Care 2. Control 3. Clarify 4. Challenge 5. Captivate 6. Confer 7. Consolidate Also includes additional **engagement** items on: ▪ Academic goals and behaviors ▪ Academic beliefs and feelings ▪ Social goals and behaviors ▪ Social beliefs and feelings	Grades 3–5/6–12: 1. No, never/Totally untrue 2. Mostly not/Mostly untrue 3. Maybe, sometimes/ Somewhat 4. Mostly yes/Mostly true 5. Yes, always/Totally true Grades K–2: No, Maybe, Yes	**Clarify:** "My teacher explains difficult things clearly." **Consolidate:** "My teacher takes the time to summarize what we learn each day." **Control:** "Our class stays busy and doesn't waste time."
YouthTruth (www.youthtruthsurvey.org) ▪ Developed and distributed by the Center for Effective Philanthropy ▪ Started as school-focused tool at secondary level; now includes classroom-level items adapted from YouthTruth Rigor & Relationships constructs and from Tripod 7 Cs highlighted by MET project study ▪ Versions for grades 6–8 and 9–12; online and paper ▪ TNTP is using YouthTruth to administer Tripod-based items as one component in evaluation to determine if TNTP-trained novice teachers are recommended for licensure ▪ Analysis for predictive validity and reliability of classroom-level items as administered by YouthTruth forthcoming	▪ Version used by TNTP includes 25 items	Constructs drawn from Tripod 7 Cs highlighted by the MET project study Plus: ▪ Rigor ▪ Relationships	1. Strongly disagree 2. Somewhat disagree 3. Neither agree nor disagree 4. Somewhat agree 5. Strongly agree	▪ **Rigor:** "The work that I do in this class really makes me think." ▪ **Relationships:** "My teacher is willing to give extra help on school work if I need it."

(continued on p.8)

(continued)

Tool and Background	Number of Items	Constructs	Response Options	Sample Items
My Student Survey (www.mystudentsurvey.com) ▪ Developed by Ryan Balch, Expert Fellow at Vanderbilt University ▪ Based on research-based teaching practices and validated observation rubrics such as Framework for Teaching ▪ Versions for grades 4–5 and 6–12 ▪ Developer-led study found results reliably consistent and predictive of student achievement and student engagement when administered to 15,000 students in Georgia (see "My Student Survey" site for research report)	▪ 55 items	1. Presenter 2. Manager 3. Counselor 4. Coach 5. Motivational Speaker 6. Content Expert	1. Never 2. Sometimes 3. Often 4. Almost always 5. Every time	**Presenter:** "When presenting new skills or ideas in class, my teacher tells us about mistakes that students often make." **Coach:** "My teacher gives us guidelines for assignments so we know how we will be graded (grading rules, charts, rubrics, etc.)."
iKnowMyClass (www.iKnowMyClass.com) ▪ Developed by Russell Quaglia at the Quaglia Institute for Student Aspirations in Portland, ME, as tool for teacher feedback ▪ Online administration only; no capacity to link individual student results to other data ▪ Focus on student engagement and relationships ▪ Classroom-level survey available for grades 6–12; version for grades 3–5 planned for release September 2012 ▪ Developer-led study validated 6–12 version for "construct validity" (items within each construct found to be related), but no study yet on tool's ability to predict student outcomes ▪ Quaglia is also working with the Pearson Foundation to adapt its school-level "MyVoice" survey for classroom level (grades 3–5 and 6–12) (see www.myvoice.pearsonfoundation.org)	▪ Full and short versions for grades 6–12 ▪ Full: 50 items ▪ Short: 20 items ▪ Grades 3–5 version to have 27 items	1. Engagement 2. Relevance 3. Relationships 4. Class Efficacy 5. Cooperative Learning Environment 6. Critical Thinking 7. Positive Pedagogy 8. Discipline Problems	1. Strongly disagree 2. Disagree 3. Undecided 4. Agree 5. Strongly agree	**Class Efficacy:** "I feel comfortable asking my teacher for individual help about the things we are learning." **Cooperative Learning:** "The teacher encourages students to work together." **Positive Pedagogy:** "I am encouraged to use my imagination."

Checking for Alignment with a Theory of Instruction

Under a new evaluation system designed with input from teacher leaders and administrators in the **Memphis City Schools**, "stakeholder perceptions"—including student survey responses—account for 5 percent of teachers' overall ratings. Other components include:

- Classroom observations: 40%
- Student achievement/learning gains: 50%
- Evidence of content knowledge: 5%

Designers of the evaluation system recognized the importance of alignment between the different components so that teachers received consistent messages about their performance. In adopting Tripod as its student survey, they mapped the instrument's constructs and items to 11 dimensions in the district's evaluation framework that relate to teaching and the classroom environment. They found all of the dimensions represented to some degree in the survey. Examples in the table below illustrate the alignment.

Alignment of Evaluation Framework and Survey Components

	Memphis City Schools Teacher Evaluation Rubric	Tripod Student Survey
Evaluation rubric dimension/survey construct	Use strategies that develop higher-level thinking skills	Challenge
Example indicator/item	Questions require students to apply, evaluate, or synthesize	"My teacher wants me to explain my answers—why I think what I think."
Evaluation rubric dimension/survey construct	Check for understanding and respond appropriately during the lesson	Clarify
Example indicator/item	If an attempt to address a misunderstanding is not succeeding, the teacher, when appropriate, responds with another way of scaffolding	"If you don't understand something, my teacher explains it another way."

Predicting Outcomes

Measuring what matters also means capturing aspects of teaching and the learning environment that relate to desired student outcomes. If not, the resulting feedback and evaluation won't support the improvements a school system cares about—or worse, it may be counterproductive. This refers to "predictive validity"—the extent to which survey results predict which teachers will have more student achievement gains. For a survey to be predictively valid, it means that, on average, the teachers who get the most favorable survey responses are also those who are helping students learn the most. If students perceive differences among teachers, those differences should generally predict student outcomes.

The MET project's analysis of Tripod found this to be true for the sample of teachers it studied. MET project researchers ranked teachers based on the proportion of their students who gave favorable responses to the items within each of the 7 Cs. When they then looked at those same teachers' student achievement gains when teaching other students, they found that those identified as being in the top 25 percent

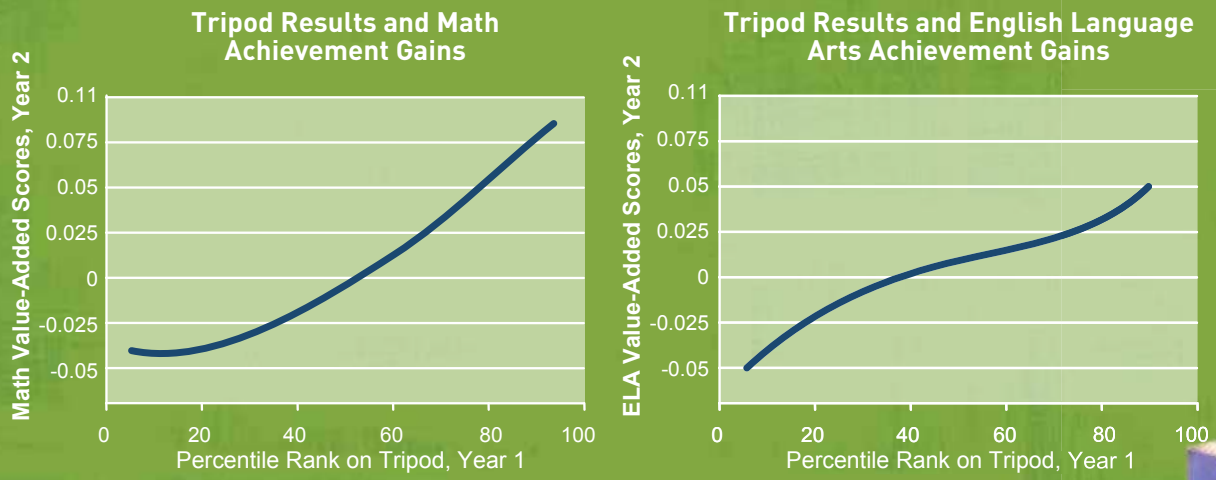
based on Tripod results had students who were learning the equivalent of about 4.6 months of schooling more in math—over the course of the year and measured using state tests—than the students of teachers whose survey results were in the bottom 25 percent. Tripod's predictive power accounts for less than half as many months difference when it came to gains based on state English language arts (ELA) tests (a common finding among many measures), but a clear relationship was found nonetheless.

The positive relationship between teachers' Tripod results and their student achievement gains is further shown in **Figure 2.**

Systems that use student surveys should similarly test for predictive validity by comparing teachers' survey results with their student achievement gains—as they should for any measure they use. Moreover, checking for predictive validity needs to be an ongoing process. Over time, alignment with student outcomes could deteriorate. This could happen if somehow teachers altered their actions in ways that improved their survey results, but without improving their underlying performance on practices associated with better outcomes. In such a situation, a system would see that teachers' rankings based on survey results bore little relationship to their students' learning gains. Such misalignment could signal the need for survey refinement.

Figure 2

Teachers' Tripod Student Survey Results Predict Achievement Gains



These slopes represent the relationship between teachers' rankings based on Tripod student survey results from one year and their "value-added" scores—a measure of student achievement gains—from the following year. As shown, teachers with more favorable Tripod results had better value-added scores. Achievement gains here are based on state assessments.

Ensure Accuracy

For results to be useful, they must be correct. It's of no benefit if survey responses reflect misunderstanding of what's being asked, or simply what students think others want them to say. Worse yet are results attributed to the wrong teacher. Honest feedback requires carefully worded items and assurances of confidentiality. Standardized procedures for how surveys are distributed, proctored, and collected are a must. Accurate attribution requires verification.

Item Wording

Survey items need to be clear to the students who respond to them. A well-crafted item asks about one thing. One wouldn't pose the statement: "This class stays busy all the time *and* the students act the way the teacher wants." Good items also avoid double-negatives, and their language is age appropriate. Wording for similar items may vary depending on the grade level. Note the slight difference in explicitness of the following two items from Tripod:

- Grades 3–5: "When my teacher marks my work, he/she writes on my papers to help me understand."
- Grades 6–12: "The comments that I get on my work in this class help me understand how to improve."

Student perception survey development involves discussion with students to determine if they're interpreting the items as intended. Through systematic "cognitive interviews," survey developers probe both students' understanding of each item and whether the item addresses its desired objective. Items that aren't clear are reworded and tested again. During survey administration, students are generally told not to answer items they don't understand as a further check against meaningless results and to indicate which items may need clarification.

Young Students and Special Populations

Well-designed surveys account for the fact that not all students read at the same grade level. Even still, school systems often involve special education teachers and teachers of English language learners (ELL) in reviewing an instrument's survey items before they are used. For some ELL students, surveys that are translated into their native language may need to be available.

Special care is also required when surveying the youngest students—some of whom are not yet able to read. Although not part of the MET project analysis, Tripod has a K–2 version of its survey, which differs from the grades 3–5 and 6–12 forms in three respects:

- It has fewer items overall;
- It has three response options—yes/no/maybe—instead of five; and

- Items are worded more simply, for example:
  - K–2: "My teacher is very good at explaining things."
  - Grades 3–5: "If you don't understand something, my teacher explains it another way."

Survey administration in grades K–2 also requires accommodations. Given the need to ensure confidentiality while also reading items to younger students, someone other than a class' teacher must proctor. Often, young students are administered surveys in small groups—not entire classes—to allow adequate support from proctors. Even with such accommodations, results should be monitored for indications that items are being understood. A lack of consistency in responses to items within the same construct, for example, might signal the need for different wording.

Ensuring Confidentiality

Confidentiality for students is a non-negotiable if surveys are part of formal feedback and evaluation. If students believe their responses will negatively influence how their teachers treat them, feel about them, or grade them, then they'll respond so as to avoid that happening. More fundamentally, students shouldn't be made to feel uncomfortable. They should be told, in words and actions, that their teachers will not know what individual students say about their classrooms.

Consistently applied protocols are essential for providing students such assurance. Although in many situations teachers will distribute student perception surveys in their own classrooms, no teacher should receive back a completed survey form that would allow the teacher to identify who filled it out. In the MET project, following procedures generally employed in administering Tripod, paper surveys were distributed to students with their names on peel-off labels that they removed before completing them. All that remained on the form when they finished were unique bar codes to let researchers link their responses to other data collected for the study but which no school personnel could use to identify respondents. Students also placed their completed forms in opaque envelopes and sealed them.

Confidentiality also requires setting a minimum number of respondents for providing teachers with results. If results for a class are based on surveys from only three or four students, then a teacher could conjecture that a highly favorable or unfavorable result for an item reflects how all of those students responded individually. (Although less likely, this still could happen if all students in a class give the same highly unfavorable or favorable response to a teacher.)

Accuracy of Attribution

Systems must be certain about which teacher and class each completed survey belongs to. Part of ensuring this requires making sure students have the right teacher and class in mind when responding, through verbal and written instructions. For classes taught by more than one teacher, systems may

Procedures for Confidentiality

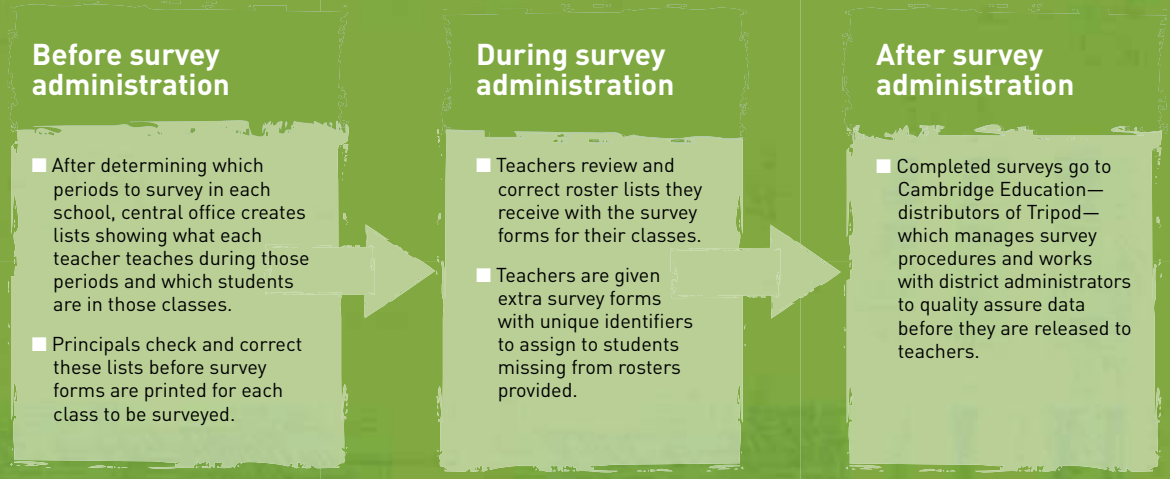
Denver Public Schools is piloting a student survey for use in a district-wide teacher evaluation system slated for full implementation in the 2014–15 school year. As part of the pilot, paper surveys were administered in about 130 schools in fall 2011 and spring 2012. One goal of the pilot is to hone administration guidelines. Here are a few ways current guidelines address confidentiality:

- **Scripts are provided to teachers on survey administration that assure students: "I will never get to see your individual student answers, only a class summary of the responses." Students are also told the surveys "will help teachers and principals understand what they do well and how they may be able to improve."**
- **Teachers assign a student the task of collecting all completed surveys, sealing them in an envelope, and taking them to the school's survey coordinator.**
- **Results from classes in which fewer than 10 students respond are not reported back to teachers because low response rates may suggest how individual students responded. Teachers are told to strive for classroom response rates of at least 80 percent. (Teachers with fewer than five students in a class do not participate.)**

Figure 3

How the Pittsburgh Public Schools Ensures Accuracy of Survey Attribution

Like many systems implementing student surveys, **Pittsburgh Public Schools** (PPS) has opted to assign unique student identifiers to each survey form to allow for more quality-control checks and more precise tracking of results. That means ensuring the accuracy of each teacher's student roster in those sections to be surveyed. This figure shows the multiple points at which PPS checks class information during each survey administration.



need to decide who is the "teacher of record"—the one primarily responsible for instruction—or administer separate surveys for each teacher in a classroom.

A significant challenge is posed by linking each student's individual responses to a particular teacher and class. Doing so while ensuring confidentiality generally requires that each survey form be assigned a unique student identifier before a student completes it. While systems can administer surveys without this feature and still attribute a classroom's group of surveys to the right teacher, they wouldn't be able to compare how the same students respond in different classrooms or at different points in time. Nor could they connect survey results with other data for the same students. Such connections can help in assessing the effectiveness of interventions and the reliability of the survey itself. They also allow for more data integrity checks.

Distributing surveys with individual identifiers for each student means generating lists of the students in each class ahead of time, and then verifying them. As simple as that may sound, determining accurate class rosters for every teacher in a district can be vexing. Until recently, few stakes have been attached to the accuracy of the student and teacher assignments recorded in the information systems from which rosters would be generated. Class lists from the start of the school year may be out of date by the time a survey is administered.

Rosters must be verified at the point in time that surveys are administered, with procedures in place to assign unique identifiers to eligible students in a class who aren't listed. (Typically, systems don't survey students unless they've been with a teacher for at least several weeks to ensure familiarity.) Many places implementing paper student surveys employ procedures similar to those used by the MET project in administering Tripod, in which each teacher receives a packet that includes:

- A class roster to check and correct if needed;
- Survey forms with unique identifiers for all students on the roster; and
- A set of extra forms that each have unique identifiers to assign eligible students in the class not listed in the roster.

Similar verification procedures were used for online administration in the MET project, but with teachers distributing login codes with unique identifiers instead of paper forms.

Ensure Reliablity

Consistency builds confidence. Teachers want to know that any assessment of their practice reflects what they typically do in their classrooms, and not some quirk derived from who does the assessing, when the assessment is done, or which students are in the class at the time. They want results to be reliable.

Capturing a Consistent View

In many ways, the very nature of student perception surveys mitigates against inconsistency. The surveys don't ask about particular points in time, but about how students perceive what typically happens in a classroom. Moreover, a teacher's results average together responses from multiple students. While one student might have a particular grudge, it's unlikely to be shared by most students in a class—or if it is, that may signal a real issue that merits addressing.

It should perhaps not be surprising then, that the MET project found Tripod to be more reliable than student achievement gains or classroom observations—the instruments studied to date. When researchers compared multiple results on the same measure for the same teachers, the student surveys were more likely than achievement gain measures or observations to demonstrate consistency. While a single observation may provide accurate and useful information on the instruction observed at a particular time on a particular day, it represents a small slice of what goes on in a classroom over the course of the year. In the context of the MET project's research, results from a single survey administration proved to be significantly more reliable than a single observation.

Numbers of Items

Although student surveys can produce relatively consistent results, reliability isn't a given. Along with the quality of the items used, reliability is in part a function of how many items are included in a survey. Both reliability and feedback can be enhanced by including multiple items for each of a survey's constructs.

Online or Paper?

Online and paper versions of the same survey have been found to be similarly reliable. Instead, the question of which form to use—or whether to allow for both—is largely a matter of capacities. Online surveys allow for more automated processing and are less prone to human error, but they require an adequate number of computers within each school and sufficient bandwidth to handle large numbers of students going online at the same time.

Adapting and Streamlining a Survey Instrument

After initially administering a version of the Tripod survey similar to that used in the MET project (with some 75 items), **Denver Public Schools** (DPS) officials heard from some teachers concerned that the survey took too long for students to complete. For this and other reasons, the system engaged teachers in deciding how to streamline a tool while still asking multiple items within each Tripod construct. The result for 2011–12: a survey that includes three questions for each of the Tripod's 7 Cs. (The district's prekindergarten through grade 2 version includes nine items total, with the construct of "Care" as the only one with multiple items.) The table below compares the items under "Care" in the upper elementary versions of the Tripod-MET project survey with those in DPS's instrument to illustrate the district's adaptation. The school system is assessing results to determine the reliability of the streamlined tool and the value teachers see in the feedback provided.

Tripod-MET project version	Denver Public Schools' adaptation
Answers on 5-point scale: 1. Totally untrue 2. Mostly untrue 3. Somewhat true 4. Mostly true 5. Totally true	Answers on 4-point scale 1. Never 2. Some of the time 3. Most of the time 4. Always
Care	
▪ I like the way my teacher treats me when I need help.	▪ My teacher is nice to me when I need help.
▪ My teacher is nice to me when I ask questions.	
▪ My teacher gives us time to explain our ideas.	
▪ If I am sad or angry, my teacher helps me feel better.	▪ If I am sad or angry, my teacher helps me feel better.
▪ My teacher seems to know when something is bothering me.	
▪ The teacher in this class encourages me to do my best.	▪ The teacher in this class encourages me to do my best.
▪ My teacher in this class makes me feel that he/she really cares about me.	

The Tripod surveys administered in the MET project included more than 75 items, but analysis for reliability and predictive validity was based on 36 items that related to the 7 Cs (see the appendix for a complete list of the 7 C items in the MET project's analysis). Other questions asked about students' backgrounds, effort, and sense of self-efficacy. To understand how different-but-related items can add informational value, consider these two under the construct of "Clarify" in the Tripod survey:

- "If you don't understand something, my teacher explains it another way."
- "My teacher has several good ways to explain each topic that we cover in class."

Both relate to a similar quality, but in different ways. A favorable response on both presumably is a stronger indicator of Clarify than a favorable response on just one. Moreover, one can imagine a teacher who often has several good ways to explain topics but who doesn't recognize when students misunderstand. Different results for somewhat different-but-related items can better pinpoint areas for improvement.

At the same time, too many items might result in a survey taking too much instructional time to complete. Such concerns lead many systems to shorten their surveys after initial piloting. In doing so, the balance they strive for is to maintain reliability and the promise of useful feedback while producing a streamlined tool. Forthcoming analyses by MET project partners will explore in greater detail the relationship between reliability and number of items.

Sampling

Even a comparatively reliable measure could be made more so by using bigger samples. When the MET project analyzed a subset of participating classrooms in which Tripod surveys were given in December and March, researchers found significant stability across the two months for the same teacher. However, averaging together results from different groups of students for the same teacher would reduce the effects of any variance due to the make-up of a particular class. MET project research partners now are analyzing results to determine the payoff in increased reliability from doing so. But in practice, many school systems are including two survey administrations per year in their pilots as they weigh the costs of the approach against the benefit of being able to gauge improvement during the year and of distributing the stakes attached to survey results across multiple points in time. Such systems often survey different sections during each administration for teachers who teach multiple sections.

Sampling for Teachers Who Teach Multiple Sections

Deciding which students to survey for a teacher who teaches a self-contained classroom is straightforward: survey the entire class. But a more considered approach is needed for teachers who teach multiple sections. Systems have experimented with varied approaches:

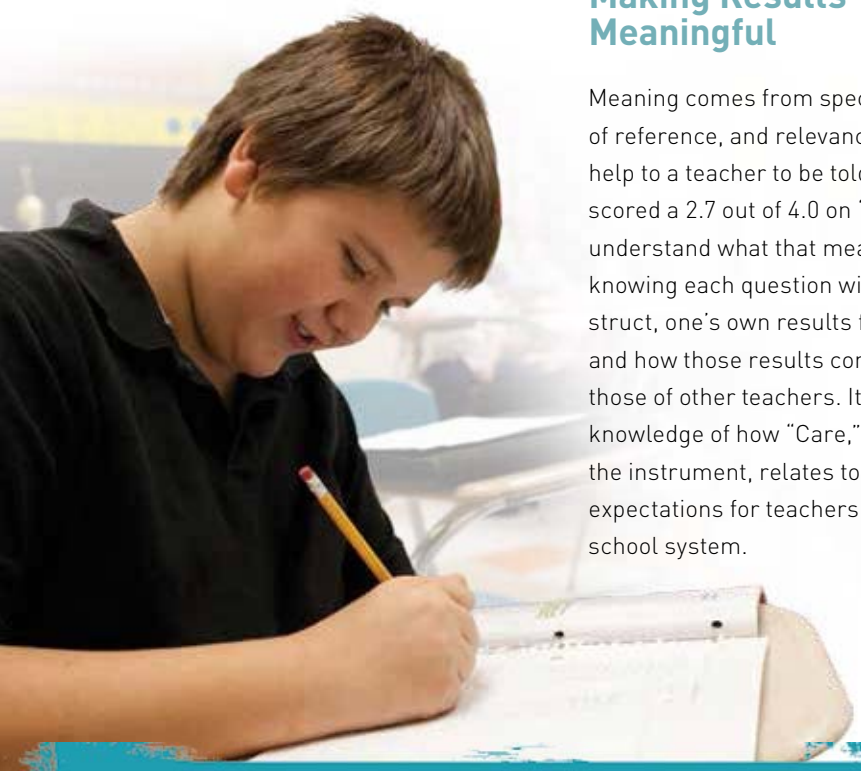
- **Survey all students on all of their teachers.** One option that requires minimal planning to implement is to survey all students on all of their teachers. But doing so would have individual students completing surveys on several teachers, which could prompt concerns of overburdening students. Aspire Public Schools has recently tried this approach but only after reducing the number of questions in its survey.

- **Survey one class period per teacher with each administration.** While reducing the potential for overburdening students, this approach still results in some students completing surveys on more than one teacher because typically in a school there is no one class period during which all teachers teach. Whenever surveys are given during more than one period, chances are some students will complete surveys on more than one teacher. Pittsburgh Public Schools, which has surveyed twice a year, determines guidelines for teachers at each school on which period to survey based on analysis of class rosters meant to minimize the burden on students. The district surveys during different periods in the fall and spring to capture the views of different students.

- **Survey a random sample across sections.** After trying various strategies, Green Dot Public Schools settled on one designed to minimize the burden on students while addressing concerns teachers expressed that one section may not be adequately representative. Twice a year, Green Dot uses software to randomly draw the names of at least 25 students from all of those taught by each teacher. It does so in such a way that no student in a school is listed for more than two teachers. Charter management organization officials use those lists to create surveys that are then administered to students in "advisory classes," a kind of homeroom period focused on study skills and youth development. So different students in the same advisory complete surveys for different teachers—but across all advisories, each teacher has at least 25 randomly selected students who are surveyed. Green Dot stresses that the strategy requires a highly accurate and up-to-date student information system.

Support Improvement

Accurate measurement is essential, but it is insufficient to improve effectiveness. Imagine the athlete whose performance is closely monitored but who never receives feedback or coaching. Likewise, without providing support for improvement, school systems will not realize a return on their investments in student surveys. That means helping teachers understand what to make of their results and the kinds of changes they can make in their in practice to improve them.



The MET project has not completed analysis on the effectiveness of feedback and coaching, instead focusing initial research on what it takes to produce the kind of valid and reliable information on which such support depends. But examples of how survey results can be used in feedback and support may be found among those school systems at the forefront of implementation.

Making Results Meaningful

Meaning comes from specificity, points of reference, and relevance. It's of little help to a teacher to be told simply "you scored a 2.7 out of 4.0 on 'Care.' " To understand what that means requires knowing each question within the construct, one's own results for each item, and how those results compare with those of other teachers. It also requires knowledge of how "Care," as defined by the instrument, relates to the overall expectations for teachers within the school system.

For these reasons, how and when results are presented to teachers is critically important if student surveys are to support improvement. Data displays should be complete and easy to interpret, allowing teachers to see the strengths and weaknesses within their classrooms and to see how the evidence supports those judgments. Ideally, they should integrate results from multiple measures, calling out the connections among results from surveys and classroom observations, student achievement gains, and other evaluation components.

The online teacher-evaluation information system developed by **Green Dot Public Schools** illustrates how survey results can be presented meaningfully. As shown in the excerpt in **Figure 4,** teachers who access their results via the platform can compare their results on each item to school-wide averages and to average results across all Green Dot schools. Overall averages are provided, as well as the distribution of responses for each item—showing the extent to which students in a class feel similarly.

Another screen on Green Dot's platform shows teachers their survey results organized by each indicator in the charter management organization's teacher evaluation rubric—for example, showing their average results for all items related to the indicator of "communicating learning objectives to students." Elsewhere a high-level dashboard display presents teachers' summary results on each of Green Dot's measures—observations, surveys, and student achievement gains—as well as overall effectiveness ratings.

Coaching

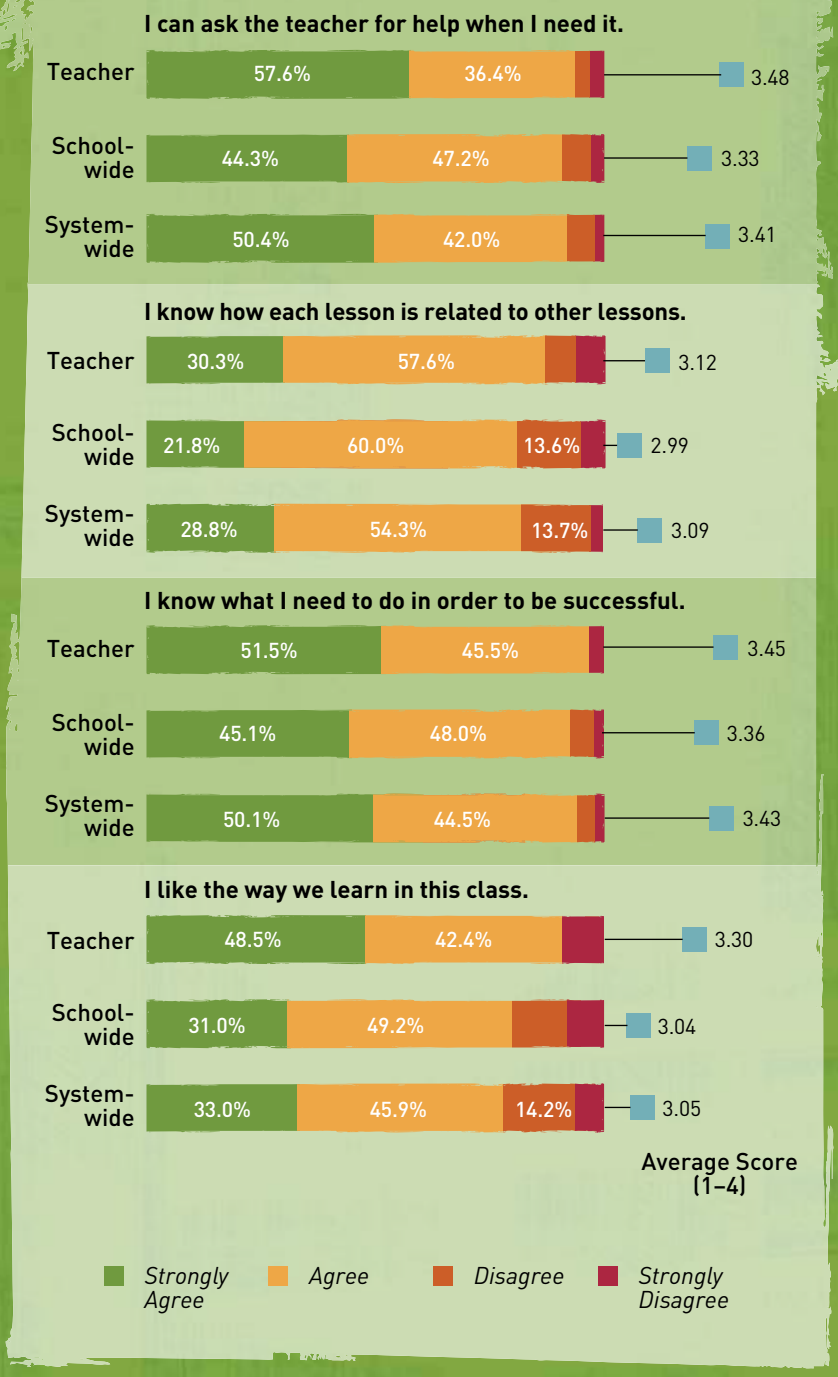
Rare are the individuals who get better at something by themselves. For most people, improvement requires the example and expertise of others. While student surveys can help point to areas for improvement, they can't answer the question: Now what? Motivation without guidance is a recipe for frustration.

Unfortunately, education has a generally poor track record when it comes to providing effective professional development. Although teachers now employ cycles of data-based instructional planning to support their students' learning, school systems have lagged far behind in implementing cycles of professional growth planning for teachers. One-shot workshops remain the norm in many places.

Even many school systems at the forefront of implementing student surveys have yet to meaningfully integrate the tools into robust coaching and planning structures. One example of what such a structure might look like comes from a research project under way in Memphis, and soon to expand to Charlotte, NC: The Tripod Professional

Figure 4

Green Dot's Display of Student Survey Results



Development (PD) research project, which includes several MET project partners: Cambridge Education (distributors of the Tripod survey), **Memphis City Schools** and the **Charlotte-Mecklenburg Schools,** and researchers from the University of Virginia, Harvard, and the University of Chicago. Funding comes from the Bill & Melinda Gates Foundation.

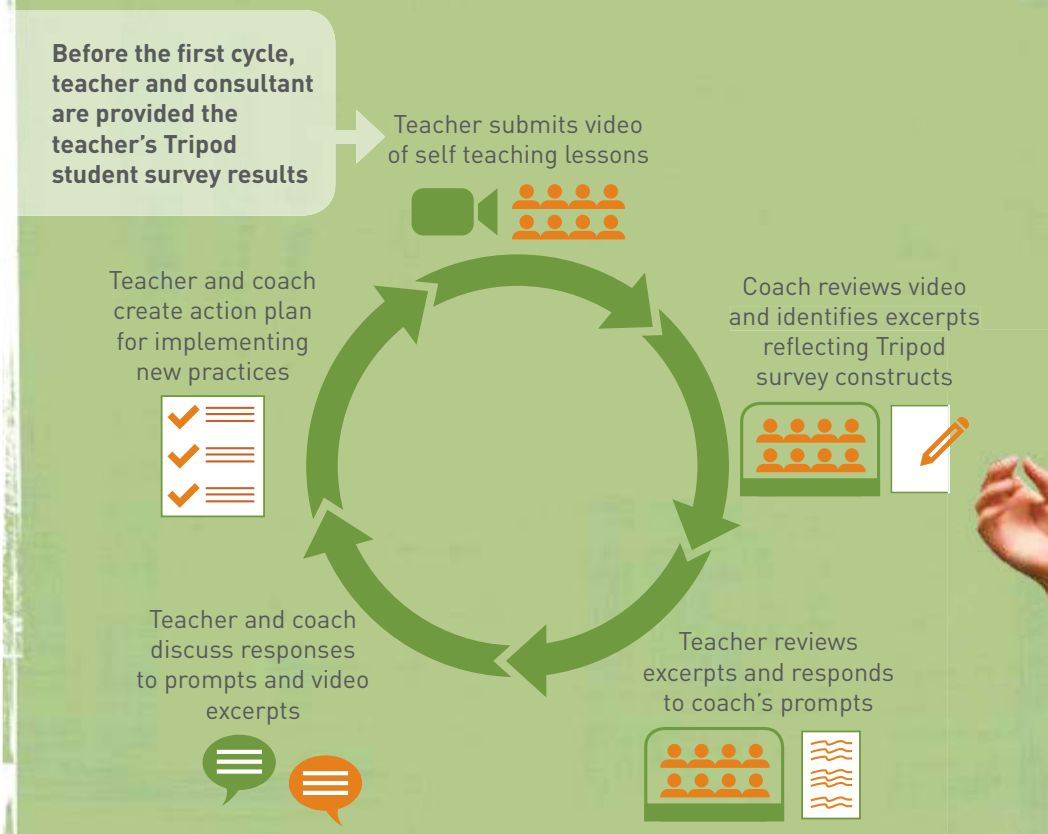
The Tripod PD research project combines the use of student surveys with a video-based one-on-one coaching cycle—called MyTeachingPartner—created at the University of Virginia by developers of the **Classroom Assessment Scoring System (CLASS),** one of several observation instruments studied by the MET project. In each cycle, teachers submit recordings of themselves engaged in teaching, and then they work with trained coaches by phone and e-mail over two to three weeks to analyze their instruction and create plans for implementing new practices in the classroom.

In the PD project, teachers and their coaches begin the first cycle informed by the teachers' Tripod survey results, and coaching protocols focus video-analysis through each cycle on the components of teaching represented in Tripod's 7 Cs. So a coach might use a one-minute excerpt from the video of the teacher's lesson that illustrates clarification to organize dialogue around how the observed practice supports student understanding. To implement action plans, teachers will have access to a video library of clips showing exemplary practice within each of Tripod's 7 Cs. Participants in the Tripod PD coaching process will go through eight cycles in a school year.

To evaluate the support's effectiveness, the research project includes a randomized experiment in which teachers who receive coaching will be compared to those who don't in terms of their student achievement gains and changes in their Tripod results. Along with one-on-one coaching, the pilot is testing the effectiveness of support provided via an online community that includes discussion forums and access to articles related to each C, as well as access to the **MyTeachingPartner** video library. Full implementation of the project is taking place during the 2012–13 school year, and results of the research are to be released in 2014.

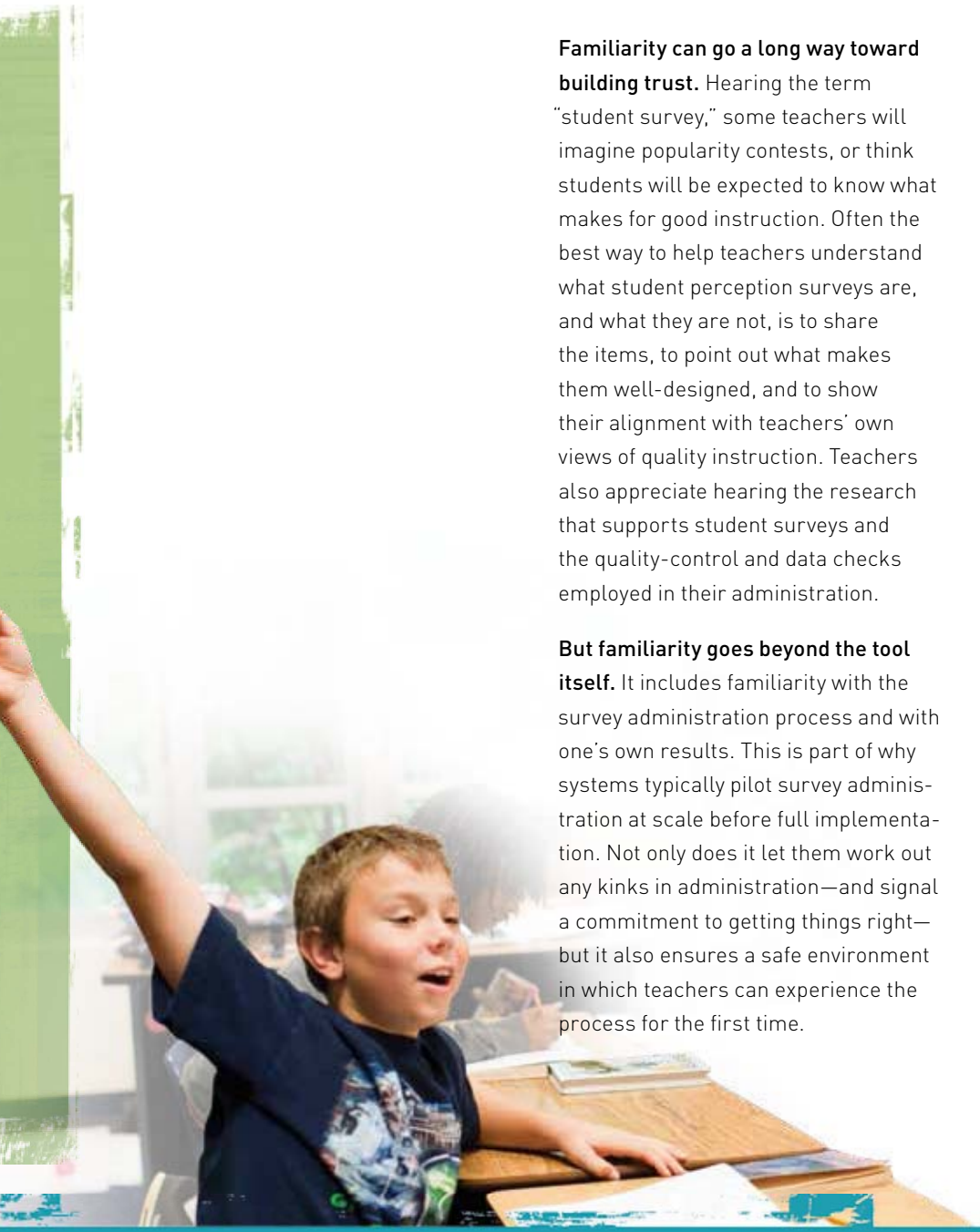
Figure 5

Memphis MyTeachingPartner-Tripod PD Coaching Cycle



Engage Stakeholders

No policy can succeed without the trust and understanding of those who implement it. Although often ill-defined, collaboration is key to effective implementation of new practices that have consequences for students and teachers. Every system referenced in this brief has made stakeholder engagement central to its rollout of student surveys—involving teachers in the review of survey items, administration procedures, and plans for using results in feedback and evaluation.



Familiarity can go a long way toward building trust. Hearing the term "student survey," some teachers will imagine popularity contests, or think students will be expected to know what makes for good instruction. Often the best way to help teachers understand what student perception surveys are, and what they are not, is to share the items, to point out what makes them well-designed, and to show their alignment with teachers' own views of quality instruction. Teachers also appreciate hearing the research that supports student surveys and the quality-control and data checks employed in their administration.

But familiarity goes beyond the tool itself. It includes familiarity with the survey administration process and with one's own results. This is part of why systems typically pilot survey administration at scale before full implementation. Not only does it let them work out any kinks in administration—and signal a commitment to getting things right—but it also ensures a safe environment in which teachers can experience the process for the first time.

Another useful engagement strategy is dialogue. Systems at the leading edge of survey implementation have opened multiple lines of two-way communication focused on their tools and procedures. They've held focus groups, polled teachers on their experiences taking part in student survey pilots, and provided clear points of contact on all survey-related issues. This increases the likelihood that people's questions will be answered, communicates that teachers' opinions matter, and can reveal problems that need fixing.

Finally, effective engagement means communicating a commitment to learn. The use of student perception surveys in feedback and evaluation is a practice in its infancy. Although much has been learned in recent years about the promise of surveys as an important source of information, many approaches toward their implementation are only now being tested in the field for the first time. Lessons learned from those experiences, combined with further research, will suggest clearer guidance than this document provides. Working together will produce better practice.

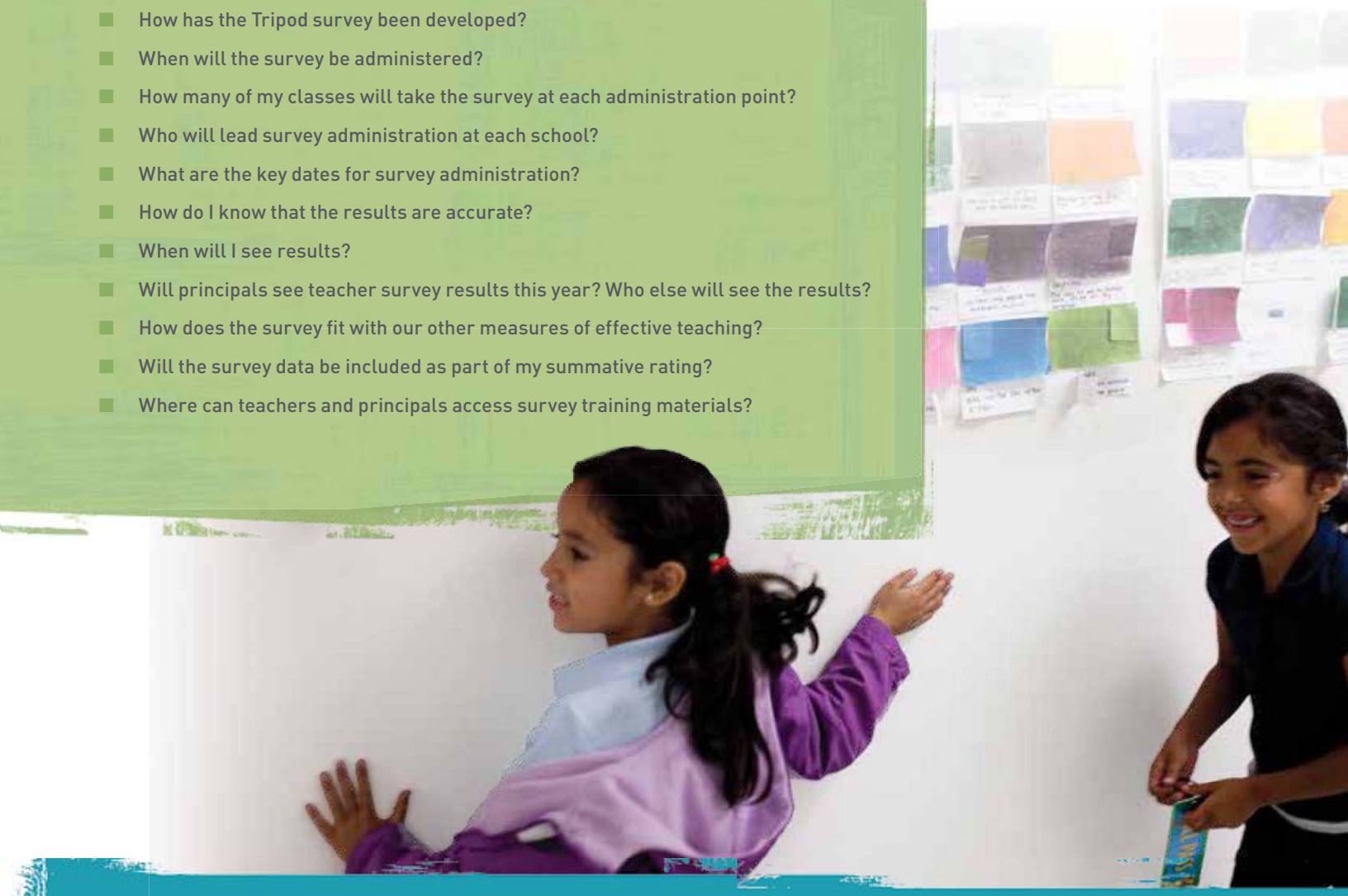
Listening and responding to teachers in Pittsburgh

To communicate with teachers on elements of the **Pittsburgh Public School's** new evaluation system—called RISE, for Research-Based Inclusive System of Evaluation—the district relies heavily on teams of informed teacher leaders at each school, called RISE leadership teams, who also participate in a district-wide committee that helps plan Pittsburgh's new measures of teaching.

In school-level conversations with teachers organized through these teams, district leaders have identified specific concerns and ideas for addressing them. For example, they learned in these discussions that teachers wanted more guidance on how to explain to students the purpose of the surveys.

In addition, system leaders have prepared a comprehensive FAQ document that addresses many of the questions teachers ask. Among the 35 questions now answered in the document:

- What is the Tripod student survey?
- What kind of questions are on the survey?
- Where can I find more information about Tripod?
- How has the Tripod survey been developed?
- When will the survey be administered?
- How many of my classes will take the survey at each administration point?
- Who will lead survey administration at each school?
- What are the key dates for survey administration?
- How do I know that the results are accurate?
- When will I see results?
- Will principals see teacher survey results this year? Who else will see the results?
- How does the survey fit with our other measures of effective teaching?
- Will the survey data be included as part of my summative rating?
- Where can teachers and principals access survey training materials?



Appendix

Tripod 7 C Survey Items Included in MET Project Analysis

Upper Elementary Version	Secondary Version
CARE	**CARE**
I like the way my teacher treats me when I need help.	My teacher in this class makes me feel s/he really cares about me.
My teacher is nice to me when I ask questions.	My teacher seems to know if something is bothering me.
My teacher in this class makes me feel that s/he really cares about me.	My teacher really tries to understand how students feel about things.
If I am sad or angry, my teacher helps me feel better.	**CONTROL**
The teacher in this class encourages me to do my best.	Student behavior in this class is under control.
My teacher seems to know if something is bothering me.	I hate the way that students behave in this class.*
My teacher gives us time to explain our ideas.	Student behavior in this class makes the teacher angry.*
CONTROL	Student behavior in this class is a problem.*
My classmates behave the way my teacher wants them to.	My classmates behave the way my teacher wants them to.
Our class stays busy and does not waste time.	Students in this class treat the teacher with respect.
Students behave so badly in this class that it slows down our learning.*	Our class stays busy and doesn't waste time.
Everybody knows what they should be doing and learning in this class.	**CLARIFY**
CLARIFY	If you don't understand something, my teacher explains it another way.
My teacher explains things in very orderly ways.	My teacher knows when the class understands, and when we do not.
In this class, we learn to correct our mistakes.	When s/he is teaching us, my teacher thinks we understand when we don't.*
My teacher explains difficult things clearly.	My teacher has several good ways to explain each topic that we cover in class.
My teacher has several good ways to explain each topic that we cover in this class.	My teacher explains difficult things clearly.
I understand what I am supposed to be learning in this class.	**CHALLENGE**
My teacher knows when the class understands, and when we do not.	My teacher asks questions to be sure we are following along when s/he is teaching.
This class is neat—everything has a place and things are easy to find.	My teacher asks students to explain more about the answers they give.
If you don't understand something, my teacher explains it another way.	In this class, my teacher accepts nothing less than our full effort.
CHALLENGE	My teacher doesn't let people give up when the work gets hard.
My teacher pushes us to think hard about things we read.	My teacher wants me to explain my answers—why I think what I think.
My teacher pushes everybody to work hard.	In this class, we learn a lot almost every day.
In this class we have to think hard about the writing we do.	In this class, we learn to correct our mistakes.
In this class, my teacher accepts nothing less than our full effort.	

(continued on p.24)

* Reverse coded item. Agreement represents an unfavorable response.

(continued)

Upper Elementary Version	Secondary Version
CAPTIVATE	**CAPTIVATE**
School work is interesting.	This class does not keep my attention—I get bored.*
We have interesting homework.	My teacher makes learning enjoyable.
Homework helps me learn.	My teacher makes lessons interesting.
School work is not very enjoyable.* (Do you agree?)	I like the way we learn in this class.
CONFER	**CONFER**
When s/he is teaching us, my teacher asks us whether we understand.	My teacher wants us to share our thoughts.
My teacher asks questions to be sure we are following along when s/he is teaching.	Students get to decide how activities are done in this class.
My teacher checks to make sure we understand what s/he is teaching us.	My teacher gives us time to explain our ideas.
My teacher tells us what we are learning and why.	Students speak up and share their ideas about class work.
My teacher wants us to share our thoughts.	My teacher respects my ideas and suggestions.
Students speak up and share their ideas about class work.	**CONSOLIDATE**
My teacher wants me to explain my answers—why I think what I think.	My teacher takes the time to summarize what we learn each day.
CONSOLIDATE	My teacher checks to make sure we understand what s/he is teaching us.
My teacher takes the time to summarize what we learn each day.	We get helpful comments to let us know what we did wrong on assignments.
When my teacher marks my work, s/he writes on my papers to help me understand.	The comments that I get on my work in this class help me understand how to improve.

* Reverse coded item. Agreement represents an unfavorable response.

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.

Share: 

BILL & MELINDA GATES *foundation*

www.gatesfoundation.org