



- SUBSCRIBE
- RENEW
- GIVE A GIFT
- DIGITAL EDITION

[Print](#) | [Close](#)

QUARTZ

There's a new global economy.
Follow it now | qz.com | [@quartznews](https://twitter.com/quartznews)

Why Kids Should Grade Teachers

A DECADE AGO, AN ECONOMIST AT HARVARD, RONALD FERGUSON, WONDERED WHAT WOULD HAPPEN IF TEACHERS WERE EVALUATED BY THE PEOPLE WHO SEE THEM EVERY DAY—THEIR STUDENTS. THE IDEA—AS SIMPLE AS IT SOUNDS, AND AS FAMILIAR AS IT IS ON COLLEGE CAMPUSES—WAS REVOLUTIONARY. AND THE RESULTS SEEMED TO BE, TOO: REMARKABLE CONSISTENCY FROM GRADE TO GRADE, AND ACROSS RACIAL DIVIDES. EVEN AMONG KINDERGARTEN STUDENTS. A GROWING NUMBER OF SCHOOL SYSTEMS ARE ADMINISTERING THE SURVEYS—AND MIGHT BE ABLE TO OVERCOME TEACHER RESISTANCE IN ORDER TO LINK RESULTS TO SALARIES AND PROMOTIONS.

By Amanda Ripley

Nubia Baptiste had spent some 665 days at her Washington, D.C., public school by the time she walked into second period on March 27, 2012. She was an authority on McKinley Technology High School. She knew which security guards to befriend and where to hide out to skip class (try the bleachers). She knew which teachers stayed late to write college recommendation letters for students; she knew which ones patrolled the halls like guards in a prison yard, barking at kids to disperse.

If someone had asked, she could have revealed things about her school that no adult could have known. Once Nubia got talking, she had plenty to say. But until that morning of her senior spring, no one had ever asked.

She sat down at her desk and pulled her long, neat dreadlocks behind her shoulders. Then her teacher passed out a form. Must be another standardized test, Nubia figured, to be finished and forgotten. She picked up her pencil. By senior year, it was a reflex. The only sound was the hum of the air conditioning.

Teachers in the hallway treat me with respect, even if they don't know me.

Well, this was different. She chose an answer from a list: *Sometimes.*

This class feels like a happy family.

She arched an eyebrow. Was this a joke? *Totally untrue.*

In towns around the country this past school year, a quarter-million students took a special survey designed to capture what they thought of their teachers and their classroom culture. Unlike the vast majority of surveys in human history, this one had been carefully field-tested. That research had shown something remarkable: if you asked kids the right questions, they could identify, with uncanny accuracy, their most—and least—effective teachers.

The point was so obvious, it was almost embarrassing. Kids stared at their teachers for hundreds of hours a year, which might explain their expertise.

MORE ON EDUCATION

A National Report Card



A visual look at the educational successes and failures of the past year

by Nicole Allan



Their survey answers, it turned out, were more reliable than any other known measure of teacher performance—including classroom observations and student test-score growth. All of which raised an uncomfortable new question: Should teachers be paid, trained, or dismissed based in part on what children say about them?

To find out, school officials in a handful of cities have been quietly trying out the survey. In D.C. this year, six schools participated in a pilot project, and *The Atlantic* was granted access to observe the four-month process from beginning to end.

At McKinley, a magnet school for science, technology, engineering, and mathematics, Nubia Baptiste filled in bubbles in response to all 127 questions. Then she slipped the survey into the envelope provided and sealed it.

Afterward, in the hallway, she tried to understand what had just happened. It didn't fit with her previous experience. "No one asks about the adults," she said. "It's always the student."

A classmate standing next to her shook her head. "They should've done this since I was in the eighth grade."

For the past decade, education reformers worldwide have been obsessed with teaching quality. Study after study has shown that it matters more than anything else in a school—and that it is too low in too many places. For all kids to learn 21st-century skills, teaching has to get better—somehow.

In the United States, the strategy has been for school officials to start evaluating teacher performance more frequently and more seriously than in the past, when their reviews were almost invariably positive. The hope was that a teacher would improve through a combination of pressure and feedback—or get replaced by someone better. By the beginning of this year, almost half the states required teacher reviews to be based in part on test-score data.

So far, this revolution has been loud but unsatisfying. Most teachers do not consider test-score data a fair measure of what students have learned. Complex algorithms that adjust for students' income and race have made test-score assessments more fair—but are widely resented, contested, or misunderstood by teachers.

Meanwhile, the whole debate remains moot in most classrooms. Despite all the testing in American schools, most teachers still do not teach the subjects or grade levels covered by mandatory standardized tests. So no test-score data exists upon which they can be judged. As a result, they still get evaluated by their principals, who visit their classrooms every so often and judge their work just as principals have always done—without much accuracy, detail, or candor. Even in Washington, D.C., which has been more aggressive than any other city in using test-score data to reward and fire teachers, such data have been collected for only 15 out of every 100 teachers. The proportion is increasing in D.C. Public Schools and other districts as schools pile on more tests, but for now, only a minority of teachers can be evaluated this way.

But even if testing data existed for everyone, how informative would they really be? Test scores can reveal when kids are not learning; they can't reveal *why*. They might make teachers relax or despair—but they can't help teachers

The Homeschool Diaries



In New York City, teaching your own kids can make the most sense.

by Paul Elie

The Writing Revolution



When New Dorp High School was faced with closure, the principal launched a dramatic new writing initiative—one that has become a model for educational reform.

by Peg Tyre

The Schoolmaster



David Coleman is a poetry-loving Rhodes Scholar and former McKinsey consultant whose pending overhaul of the SAT has reignited a national debate over how much we should expect from students and schools.

by Dana Goldstein



improve. The surveys focus on the means, not the ends—giving teachers tangible ideas about what they can fix right now, straight from the minds of the people who sit in front of them all day long.

A decade ago, a Harvard economist named Ronald Ferguson went to Ohio to help a small school district figure out why black kids did worse on tests than white kids. He did all kinds of things to analyze the schoolchildren in Shaker Heights, a Cleveland suburb. Maybe because he'd grown up in the area, or maybe because he is African American himself, he suspected that important forces were at work in the classroom that teachers could not see.

So eventually Ferguson gave the kids in Shaker Heights a survey—not about their entire school, but about their specific classrooms. The results were counterintuitive. The same group of kids answered differently from one classroom to the next, but the differences didn't have as much to do with race as he'd expected; in fact, black students and white students largely agreed.

The variance had to do with the teachers. In one classroom, kids said they worked hard, paid attention, and corrected their mistakes; they liked being there, and they believed that the teacher cared about them. In the next classroom, the very same kids reported that the teacher had trouble explaining things and didn't notice when students failed to understand a lesson.

“We knew the relationships that teachers build with students were important,” says Mark Freeman, superintendent of the Shaker Heights City School District. “But seeing proof of it in the survey results made a big difference. We found the results to be exceptionally helpful.”

Back at Harvard, no one took much notice of Ferguson's survey. “When I would try to talk about it to my researcher colleagues, they were not interested,” he says, laughing. “People would just change the subject.”

Then, in 2009, the Bill & Melinda Gates Foundation launched a massive project to study 3,000 teachers in seven cities and learn what made them effective—or ineffective. Thomas Kane, a colleague of Ferguson's, led the sprawling study, called the “Measures of Effective Teaching” project. He and his fellow researchers set up many elaborate instruments to gauge effectiveness, including statistical regressions that tracked changes in students' test scores over time and panoramic video cameras that captured thousands of hours of classroom activity.

But Kane also wanted to include student perceptions. So he thought of Ferguson's survey, which he'd heard about at Harvard. With Ferguson's help, Kane and his colleagues gave an abbreviated version of the survey to the tens of thousands of students in the research study—and compared the results with test scores and other measures of effectiveness. The responses did indeed help predict which classes would have the most test-score improvement at the end of the year. In math, for example, the teachers rated most highly by students delivered the equivalent of about six more months of learning than teachers with the lowest ratings. (By comparison, teachers who get a master's degree—one of the few ways to earn a pay raise in most schools—delivered about one more month of learning per year than teachers without one.)

Students were better than trained adult observers at evaluating teachers. This wasn't because they were smarter but because they had months to form an opinion, as opposed to 30 minutes. And there were dozens of them, as opposed to a single principal. Even if one kid had a grudge against a teacher or just blew off the survey, his response alone couldn't sway the average.

“There are some students, knuckleheads who will just mess the survey up and not take it seriously,” Ferguson says, “but they are very rare.” Students who don't read the questions might give the same response to every item. But when Ferguson recently examined 199,000 surveys, he found that less than one-half of 1 percent of students did so in the first 10 questions. Kids, he believes, find the questions interesting, so they tend to pay attention. And the “right” answer is not always apparent, so even kids who want to skew the results would not necessarily know how to do it.

Even young children can evaluate their teachers with relative accuracy, to Kane's surprise. In fact, the only thing that the researchers found to better predict a teacher's test-score gains was ... past test-score gains. But in addition to being loathed by teachers, those data are fickle. A teacher could be ranked as highly effective one year according to students' test gains and as ineffective the next, partly because of changes in class makeup that have little to do with



her own performance—say, getting assigned the school’s two biggest hooligans or meanest mean girls.

Student surveys, on the other hand, are far less volatile. Kids’ answers for a given teacher remained similar, Ferguson found, from class to class and from fall to spring. And more important, the questions led to revelations that test scores did not: *Above and beyond academic skills, what was it really like to spend a year in this classroom? Did you work harder in this classroom than you did anywhere else?* The answers to these questions matter to a student for years to come, long after she forgets the quadratic equation.

The survey did not ask *Do you like your teacher? Is your teacher nice?* This wasn’t a popularity contest. The survey mostly asked questions about what students saw, day in and day out.

Of the 36 items included in the Gates Foundation study, the five that most correlated with student learning were very straightforward:

1. *Students in this class treat the teacher with respect.*
2. *My classmates behave the way my teacher wants them to.*
3. *Our class stays busy and doesn’t waste time.*
4. *In this class, we learn a lot almost every day.*
5. *In this class, we learn to correct our mistakes.*

When Ferguson and Kane shared these five statements at conferences, teachers were surprised. They had typically thought it most important to care about kids, but what mattered more, according to the study, was whether teachers had control over the classroom *and* made it a challenging place to be. As most of us remember from our own school days, those two conditions did not always coexist: some teachers had high levels of control, but low levels of rigor.

After the initial Gates findings came out, in 2010, Ferguson’s survey gained statistical credibility. By then, the day-to-day work had been taken over by Cambridge Education, a for-profit consulting firm that helped school districts administer and analyze the survey. (Ferguson continues to receive a percentage of the profits from survey work.)

Suddenly, dozens of school districts wanted to try out the survey, either through Cambridge or on their own—partly because of federal incentives to evaluate teachers more rigorously, using multiple metrics. This past school year, Memphis became the first school system in the country to tie survey results to teachers’ annual reviews; surveys counted for 5 percent of a teacher’s evaluation. And that proportion may go up in the future. (Another 35 percent of the evaluation was tied to how much students’ test scores rose or fell, and 40 percent to classroom observations.) At the end of the year, some Memphis teachers were dismissed for low evaluation scores—but less than 2 percent of the faculty.

The New Teacher Project, a national nonprofit based in Brooklyn that recruits and trains new teachers, last school year used student surveys to evaluate 460 of its 1,006 teachers. “The advent of student feedback in teacher evaluations is among the most significant developments for education reform in the last decade,” says Timothy Daly, the organization’s president and a former teacher.

In Pittsburgh, all students took the survey last school year. The teachers union objects to any attempt to use the results in performance reviews, but education officials may do so anyway in the not-too-distant future. In Georgia, principals will consider student survey responses when they evaluate teachers this school year. In Chicago, starting in the fall of 2013, student survey results will count for 10 percent of a teacher’s evaluation.

No one knows whether the survey data will become less reliable as the stakes rise. (Memphis schools are currently studying their surveys to check for such distortions, with results expected later this year.) Kane thinks surveys should count for 20 to 30 percent of a teacher’s evaluations—enough for teachers and principals to take them seriously, but not enough to motivate teachers to pander to students or to cheat by, say, pressuring students to answer in a certain way.



Ferguson, for his part, is torn. He is wary of forcing anything on teachers—but he laments how rarely schools that try the surveys use the results in a systematic way to help teachers improve. On average over the past decade, only a third of teachers even clicked on the link sent to their e-mail inboxes to see the results. Presumably, more would click if the results affected their pay. For now, Ferguson urges schools to conduct the survey multiple times before making it count toward performance reviews.

As it happens, both Kane and Ferguson, like most university professors, are evaluated partly on student surveys. Their students' opinions factor into salary discussions and promotion reviews, and those opinions are available to anyone enrolled in the schools where they teach. "I think most of my colleagues take it seriously—because the institution does," Ferguson says. "Your desire not to be embarrassed definitely makes you pay attention."

Still, Ferguson dreads reading those course evaluations. The scrutiny makes him uncomfortable, he admits, even though it can be helpful. Last year, one student suggested that he use a PowerPoint presentation so that he didn't waste time writing material on the board. He took the advice, and it worked well. Some opinions, he flat-out ignores. "They say you didn't talk about something," he says, "and you know you talked about it 10 times."

In fact, the best evidence for—and against—student surveys comes from their long history in universities. Decades of research indicate that the surveys are only as valuable as the questions they include, the care with which they are administered—and the professors' reactions to them. Some studies have shown that students do indeed learn more in classes whose instructors get higher ratings; others have shown that professors inflate grades to get good reviews. So far, grades don't seem to significantly influence responses to Ferguson's survey: students who receive A's rate teachers only about 10 percent higher than D students do, on average.

The most refreshing aspect of Ferguson's survey might be that the results don't change dramatically depending on students' race or income. That is not the case with test data: nationwide, scores reliably rise (to varying degrees) depending on how white and affluent a school is. With surveys, the only effect of income may be the opposite one: Some evidence shows that kids with the most-educated parents give slightly *lower* scores to their teachers than their classmates do. Students' expectations seemingly rise along with their family income (a phenomenon also seen in patient surveys in the health-care field). But overall, even in very diverse classes, kids tend to agree about what they see happening day after day.

In a kindergarten classroom a mile from the U.S. Capitol, Gerod, 5, is evaluating his teacher. He sits at a low table in a squat chair, his yellow school-uniform shirt buttoned all the way up, and picks up a thick red pencil.

"The first question says *This class is a happy place for me to be*," the teacher says. For very young children, Ferguson's survey includes slightly different questions, which teachers from other classrooms read aloud to kids in small groups. Gerod's usual teacher was in a neighboring classroom, so that she wouldn't influence the results.

"My answer is *No*," Gerod declares, smiling. His bright-white sneakers are swinging back and forth. The other four students in his group mark *Yes*. "This is pretty easy," one of Gerod's classmates announces.

"*Sometimes I get into trouble at school*," the teacher says.

"I say *Yes*," Gerod says.

A teacher's aide chastises him from a neighboring table: "You don't have to discuss it," she says in a loud, irritated voice. "Put an answer!" But none of the kids can seem to help themselves; after each question, they continue to announce their answers loudly and clearly.

"*Some kids learn things a lot faster than I do*."

"*Yes*," Gerod says, filling in his answer.

"*I like the things that we are learning in this class*."

Gerod is getting restless. "It's time for lunch! Almost?"

It is hard to believe that Gerod's survey would pass scientific scrutiny: a few of the statements are poorly worded for



his age level, and the whole thing is far too long. But Ferguson insists that, statistically speaking, kindergartners' judgments of teachers are quite reliable; in thousands of surveys, kids in the same kindergarten class have tended to agree with each other about their teachers.

Finally, after half an hour of this, the teacher reaches the demographic questions at the end of the survey: "*Does your family speak English at home?*"

"*Never,*" Gerod says with confidence.

"Are you sure, Gerod? English—the language we are speaking now." He changes his answer to *Yes*.

"Race or ethnicity?"

"*White,*" Gerod says, marking his answer. He is black.

Patricia Wilkins, Gerod's kindergarten teacher at Tyler Elementary School, received her survey results about two months later. She'd been teaching at the school for more than a decade, and had seen a lot of reforms come and go. She'd worked for five different principals, she said, if you included the one who was led away in handcuffs.

But she was curious about the survey results. Unlike half the teachers in D.C.'s pilot project, she clicked on the link to see her students' opinions. As she looked at the data in a small conference room during a planning period, she was quiet. Then she smiled. "I'm highest on Care. That's what I felt, but I didn't know that they felt it."

Nine out of 10 of her students said they liked the way their teacher treated them when they needed help; that was high compared with the average response from kindergartners nationwide. Her students seemed to think she challenged them, too, which was reassuring. Still, only half said their classmates stayed busy and didn't waste time. "This is very helpful," she said, nodding.

Across town, at McKinley High School, Nubia Baptiste didn't hear about the survey again that school year. That summer, her teacher, Lashunda Reynolds, read the survey results for her students and found them to be fair. "Overall, I think that the survey is a good reflection tool for teachers," she said. Still, she worried that some students might be biased for or against her, and for that reason, she would not want the results to influence her formal evaluation.

Principals can be biased, too. So can tests, as Reynolds knows. But like many other teachers, she seemed fatigued by the years of one "reform" after another—and wary of any addition to the already long list of ways she would be judged.

Nathan Saunders, the head of D.C.'s teachers union, did not seem to know much about the survey when I spoke with him about it in June. But he insisted that the results should never be used for high-stakes evaluation: "This is seen by many members of our union as just another way to vilify teachers."

Guillaume Gendre, one of Nubia Baptiste's assistant principals, saw the survey results differently. "It's very, very precious data for me," he said. For this pilot, he was not able to see teachers' names beside their results, to protect their anonymity; but he said he still found the information more useful than what standardized tests provided.

Overall, the teachers scored about average compared with their counterparts in high schools nationwide. But the variation within the school was staggering—as it is in many places. In the categories of Control and Challenge—the areas that matter most to student learning—Nubia and her classmates gave different teachers wildly different reviews. For Control, which reflects how busy and well-behaved students are in a given classroom, teachers' scores ranged from 16 to 90 percent favorable; for Challenge, the range stretched from 18 to 88 percent. Some teachers were clearly respected for their ability to explain complex material or keep students on task, while others seemed to be boring their students to death.

The results helped Gendre understand why eight in 10 students who took Advanced Placement tests at McKinley, a magnet school, didn't pass. In response to one survey item—*My teacher doesn't let people give up when the work gets hard*—fewer than a third of McKinley's students answered *Totally agree*. "This building needs to be more challenging academically, and students need to feel more valued and appreciated," Gendre concluded, staring at a printout of the results in his office during the last week of school.



This school year, Washington, D.C., will make the survey available to all principals and teachers who want to use it. Chancellor Kaya Henderson says that next year, the survey may count toward teacher pay and firing decisions. But for now, she wants to proceed with caution, after years of turbulent changes in D.C. schools. “You gotta do it right,” she says. “Otherwise, it will torpedo our chances of doing it again.”

The shorter version of the survey, used in the Gates study, is available for public use, and it would cost less than \$5 per student to implement. That is a remarkable bargain. D.C.’s standardized tests and the detailed analysis of the results cost more than \$35 per pupil tested; employing professionals to watch classes and give teachers feedback multiple times a year costs about \$97 per student.

But most districts are far too invested in test-score analysis to turn back now. The ones who do adopt student surveys will almost certainly add them to test data and classroom observations, to create a more balanced (and still more complicated) measure of teacher performance.

When I called Nubia Baptiste over the summer with the survey results, she was not surprised. “Everybody knows the good teachers from the ones who don’t really want to be in the job,” she said. When I started describing the huge variation between teachers, she interrupted me. “I lived the dynamic,” she said.

Nubia was on her way to Temple University, where she was considering studying science or engineering. Having personally witnessed many of the recent reforms in D.C., she was wise to what mattered most.

“I don’t care about the results,” she said. “I care about the change the results bring. If I come back in five years and some crappy teacher is still sitting at that crappy desk, then what was the point of the survey?”

This article available online at:

<http://www.theatlantic.com/magazine/archive/2012/10/why-kids-should-grade-teachers/309088/>

Copyright © 2013 by The Atlantic Monthly Group. All Rights Reserved.

