

FAIR TO EVERYONE:

Building the Balanced Teacher Evaluations that Educators and Students Deserve



The Education Trust

TO THE POINT

- ▶ Evaluation systems that center on student learning are better for students and teachers.
- ▶ Teacher evaluation is, first and foremost, a career development tool and a way to lift quality across the profession.
- ▶ When measuring the effectiveness of teachers, start with student learning and include multiple measures.

“[Being evaluated] was a little scary, but then I sat with my chairperson and we discussed what I did. I [thought], ‘She’s going to tell me everything I did wrong.’ But it wasn’t like I thought it would be.

It was: “Look, here’s what you’re doing that’s great, and here’s where we can improve. Let’s talk about it.”

— Teacher, Elmont, N.Y., on her school’s comprehensive approach to evaluation

EXECUTIVE SUMMARY

In schools across America, teachers know who among their peers is doing the best work and who is not. Yet our evaluation systems tend to foster the notion that all teachers perform the same way, with the same results for students. Indeed, in an attempt at equality — uniform treatment for everyone — current evaluation systems often end up being fair to no one.

Ideally, performance evaluations should serve to help teachers identify strengths and areas for development, as they work to improve their practice. Systems that work have the goal of lifting quality across the profession, aiding all teachers to become good and prompting good teachers to become great.

This paper highlights key elements of evaluations that live up to these aspirations. Quality evaluation systems include regular classroom observations by trained evaluators with clear standards. They also include measurements that consider the contribution each teacher makes to student learning over a year's time, taking into account the achievement level and remediation needs students bring to the classroom.

Ultimately, everyone stands to gain when teacher evaluation systems are designed to gauge teacher performance fairly, clearly, and comprehensively, with an eye to the kind of professional growth that fuels student learning. We hope this paper demystifies some of the newer approaches to evaluation for districts and states that might be considering them. Our aim is to illustrate why these new systems are better for teachers and students.

Fair to Everyone: Building the Balanced Teacher Evaluations that Educators and Students Deserve

BY SARAH ALMY

Melissa is an excellent teacher. She demands a lot from her seventh-grade students, 90 percent of whom come from high-poverty backgrounds, and supports them when they struggle. She plans and delivers engaging lessons involving rigorous assignments tightly aligned with state standards. She analyzes and reanalyzes student data, constantly identifying which students are making gains, and which need more attention.

And most important, she gets results. Year after year, her students make one-and-one-half to two years of gains between the beginning and the end of the year.

Down the hall is another teacher doing little of what makes Melissa's class, and her students, so successful. This teacher arrives to school with no clear plan for what he wants to teach and what he wants students to learn, instead making up lessons on the fly. When unimpressive achievement results arrive each year, he blames "these students." Unlike Melissa, he rarely analyzes the data. But if he did, he would see that over the course of a year in his class, many students make less than a year's worth of academic progress.

Melissa and her colleague's practice and impact couldn't be more different. Yet you wouldn't know it by examining their respective evaluations. For each of the last four years, both teachers received "satisfactory" ratings in their summative evaluations. Even more frustrating for Melissa, she did not receive a single item of feedback from her principal as part of the evaluation. The evaluation doesn't help her to become a better teacher, and it doesn't distinguish between her and her colleague. To Melissa, it is a meaningless exercise. For both teachers it is a squandered opportunity.

Nothing about this scenario seems fair in any way. Unfortunately, this is the reality of nearly all teacher performance evaluation systems in our country today. While teachers in most schools have a sense of who is doing good work and who is not, our evaluation systems promote the fiction that all teachers perform identically.

We all know some of the reasons for this. In most evaluation systems across the United States, we judge teachers based on a small number of fleeting classroom observations. These evaluations do little to help teachers understand the impact they have upon the learning of their students. Moreover, because so many evaluations lack detail and clarity, they provide teachers little information about what to improve or how to get better.

What began as an effort to treat everyone in the profession as equals has calcified into a system that is anything but fair to teachers. Teachers who are committed to their students' achievement are forced to make up the ground students didn't cover with previous teachers who are less committed — for the same amount of pay and little recognition. Teachers who want to get better and to develop their professional craft are provided little personalized guidance.

And these evaluation systems, designed decades ago to ensure teachers are treated as professionals, have accomplished exactly the opposite. Not only is this unfair to teachers — it is profoundly unfair to students. Our failure to distinguish among teachers allows us to avoid confronting the fundamental fallacy of suggesting that one teacher is as good as the next. This particularly shortchanges the low-income students and students of color who, year after year, are saddled with less qualified and less effective teachers.

Teachers have the power to change a student's life trajectory; the work they do matters that much. A series of strong teachers can eliminate the achievement gap between white students and students of color, leveling the playing field for all students.

But current performance evaluation systems hinder the profession's ability to maximize its impact. Teachers don't have good ways of knowing whether they are narrowing or widening their students' learning

Sarah Almy is the Director of Teacher Quality at The Education Trust.

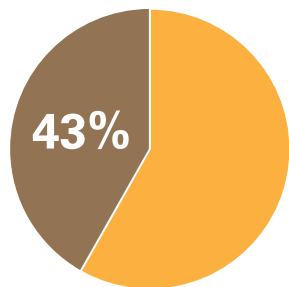
gaps. School leaders don't have good information that would enable smart decision making about how best to support and tap the talents of the teachers in their building. And students pay the price, because nothing about current systems is focused on creating better learning or ensuring stronger student outcomes.

We talk about the importance of holding our students to high expectations and recognizing the different needs of different students, yet we tolerate evaluation systems that have no real expectations for teachers and treat them all the same. These systems are failing both teachers and the students they teach.

THE PURPOSE OF PERFORMANCE EVALUATIONS

Performance evaluation is first and foremost a development tool for all teachers. It should have as its primary purpose identifying strengths and areas for growth in order to improve practice, whether a teacher is in her first year or her 14th. Evaluation systems work best when they seek to enhance the skills of every practitioner. Systems that solely aim to identify the highest and lowest performing teachers will never effectively move the needle on student achievement. These systems must be designed with the chief goals of helping all teachers to become good, and of pushing good teachers to become great.

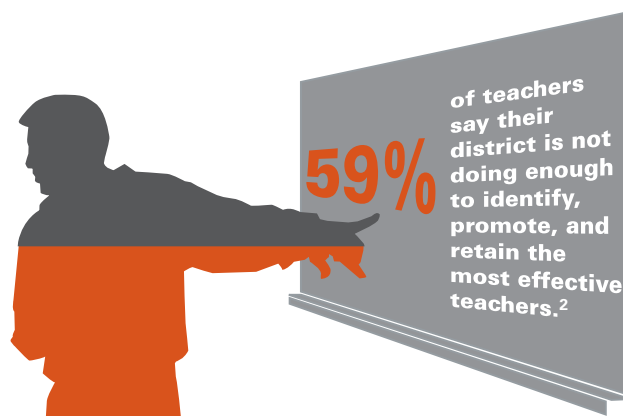
That said, using evaluation systems to address teachers at both the top and bottom of the quality spectrum is important, too. By finally identifying the true high-



Only 43 percent of teachers agree that current evaluation helps teachers improve.¹

flyers, evaluations can spotlight those teachers from whom others can and should learn. They can facilitate long overdue recognition of our strongest teachers, those who consistently achieve great things with their students, while also helping us to single out proven best practices. As well, evaluations provide a fair way to distribute incentives designed to keep top teachers where they are — or to encourage them to move where they are most needed.

Just as identifying our best teachers can help the entire profession, so can isolating that small percentage of teachers who fail our kids. Identifying and addressing these consistently poor performers will make things more fair for students who will no longer be held back by weak teachers, and for hard-working, effective



teachers who will no longer have to pick up the slack for colleagues unwilling or unable to meet the expectations of their jobs.

FOUNDATIONS FOR FAIR EVALUATIONS

Improved performance evaluation systems will not look identical across districts or states, and they shouldn't. Yet good systems will share certain critical features: classroom observations and measures of teacher impact on student learning.

Observations:

A Window into the Classroom

Most current evaluations rely exclusively on a short observation, in which administrators sit in on individual lessons caught out of context. Too often, observations are stressful experiences for teachers, feared "gotcha" moments when people who almost never see what routinely happens in their classrooms suddenly make sweeping judgments about their abilities. And in many places, observations are based on a set of vague criteria not clearly connected to student learning. In addition, though observations offer an excellent opportunity for follow-up feedback that can help teachers improve, most current systems place no priority on this. Teachers see observations as a checklist, an isolated exercise in process rather than any opportunity for meaningful feedback or professional growth.

So why would improved observations, under new evaluation systems, be better for all teachers?

Instead of a once-a-year pop-in, the new observations would form one part of a regular cycle of feedback. They would be conducted by well-trained observers who take their roles as evaluators seriously. An observation tool or rubric that is detailed and outlines clear performance standards would help establish a common language for instructional practice across schools

and districts. This tool could either be one that is developed by a district or state, or an existing high-quality tool that the district or state chooses to adopt. The important thing is that, by bringing specificity and clarity to the observation experience, a rigorous rubric keeps observations from feeling arbitrary or wildly inconsistent in the eyes of the teacher being observed. Perspectives of multiple observers — an administrator as well as someone else trained to conduct observations — and multiple observations ensure that one poorly executed lesson doesn't depict a whole year of work. Finally, observations lead to useful feedback about instruction and classroom management that result in clear, measurable goals for the teacher.

Observations can be the very best window that school leaders and others have into the teacher's mastery of her craft — her ability to manage her classroom, engage her students, deliver rigorous and accurate instruction, and gauge student understanding. More important, ongoing observations provide a teacher with thoughtful, constructive feedback that drives a cycle of learning and growth. Students deserve classroom teachers who are aware of their own strengths and weaknesses so that they can continue to improve their instruction. Real observation as part of a strong evaluation system provides this for teachers and for their students.

Measures of Student Learning: The Core of What Teachers Do

Observations provide important insight about observable practices, but they can't measure the effectiveness of those practices in terms of student outcomes. Measuring a teacher's impact on student learning lets a teacher and her supervisor know whether what she is doing in her classroom day to day is working.

Until recently, the idea of including a measure of student learning as part of a teacher's evaluation was uncommon in most places. And yet, fostering student learning is the essence of what teachers do. The overarching goal for a teacher — no matter what grade, subject area, or group of kids — is to ensure that the kids learn more than they came into the class knowing. To commend a teacher year after year for colorful bulletin boards and creative lessons while none of her students make meaningful progress over the year isn't fair — to her students, to her colleagues, or to the teacher herself who consequently misses an opportunity to improve what matters most.

Even though student achievement isn't widely included as part of existing evaluation systems, there are, of course, some principals and district leaders who look at year-end test data and make judgments about which teachers are high performers and which aren't. But making judgments based on simple achievement data — whether students made AYP or how many met the 'proficiency' standard — only tells part of the story. In most places, the data on student achievement takes neither student growth nor students' past performance into account. Teachers who spend the year doing remediation with students who entered the year already behind get no credit for the progress they make. And teachers who inherit classrooms of students already ahead look good regardless of what they do in their classrooms. Using these simple achievement results to make decisions about teachers' impact strikes many as simply unfair.

It is a good thing that no credible advocate for improving evaluation is proposing this.

Instead, advocates and policymakers have urged considering the achievement level of a class of students at the start of the school year and then measuring the contribution individual teachers make to student learning during that year.

VALUE-ADDED: WHAT DOES IT ACTUALLY MEAN?

When people talk about gauging a teacher's impact on student learning, it is most commonly referred to as a value-added measurement. Educational jargon aside, value-added is a way of measuring a student against himself and against his student peers. The calculation looks at a student in a teacher's class and then looks at how similar students — those with comparable performance histories and background characteristics — have performed in the past. Based on this information, the model predicts how the student should perform on an assessment, and then compares the predicted result with the actual result.

With this kind of data, teachers aren't punished for students who come into their classroom far below grade level, as long as the students make progress consistent with what they and other similar students have demonstrated that they have been able to do in the past. Also in this way, growth made by the initially high-achieving children of well-educated parents in one teacher's classroom is compared with the growth of other initially high achievers — not with the growth

HOW VALUE-ADDED WORKS

Andrew teaches fifth-grade reading in a school district that recently incorporated a value-added measure as part of the teacher evaluation process.



In Andrew's fifth-grade classroom this year, 90 percent of students qualify for free and reduced-price lunch, 15 percent have an IEP, and 30 percent are identified as Limited English Proficient. At the beginning of the school year, a diagnostic assessment shows Andrew that the majority of his students read at the second-grade level.

Andrew works tirelessly over the course of the school year to change his students' academic trajectory. Many of his students make tremendous progress, and he hopes that this is reflected when they take the state assessment at the end of April.

To calculate Andrew's value-added results for the current school year, the district uses multiple years of test score data for his students — in addition to other school and student factors — and figures out how much growth is typical on the fifth-grade state standardized reading exam for a group of students with similar characteristics. Using this process, the district determines that typical growth is eight points.

The district then looks at the average growth of Andrew's class. Although not all of his students have reached proficiency on the state exam, their average growth is 15 points — seven points higher than the typical growth for a group of students with similar characteristics.

Andrew's students have outpaced their predicted growth, and the amount by which they outpaced their growth is Andrew's value-added score. This information is combined with other evaluation measures as part of his overall evaluation and Andrew is identified as an effective teacher.

Measuring a teacher's impact on student learning lets a teacher and her supervisor know whether what she is doing in her classroom day to day is working.

of children in a high-poverty school who may start off at much lower levels.

The basic concept resembles a growth chart at a pediatrician's office — every child does not grow at exactly the same rate, but the doctors always want to make sure that a child is progressing along the growth chart relative to her past growth, and to how other children similar in size have progressed.

Value-added analysis is a complex concept and the equations used to determine teachers' value-added scores are difficult for many to understand. Yet this complexity also makes it a robust and fair way to accu-

Most teachers think that “how much your students are learning compared with students in other schools” is a good indication of success as a teacher.³



rately isolate a teacher's impact on a student, something previous measures have failed to do.

But what if the assessments taken by students measure low-level skills, as far too many current assessments unfortunately do? Teachers worry that these assessments, and the resulting value-added measures, will fail to capture the higher order skills that they impart to their students, and will drive teachers to teach at low levels to raise their value-added scores. However, some promising research suggests otherwise. A study of the Measures of Effective Teaching Project (MET), which spans several states and school districts, indicates that teachers with high value-added results on their students' state tests generally had high value-added scores when students were tested on higher order concepts and skills as well.⁴

Moreover, research from the same study found that teachers whose students report spending a lot of time practicing for the state test rarely show the highest value-added on state tests. The bottom line: Good teaching will shine through, whether the test is considered to be one of high quality or not.

Value-added measures are not perfect; no single performance measure ever is. However, it is essential to include some measure of whether teachers are meeting their most basic responsibility of developing student learning as part of their performance evaluation. And right now, value-added measures are by far the most equitable option we have.

BEYOND VALUE-ADDED

Despite the lively national discussion about value-added measurement and its use in evaluation systems, the reality is that in most places, these measures will only be available immediately for about one-third of teachers, at best. For those who aren't teaching subjects and grades assessed by state tests, a statewide measure of value-added results does not yet exist. But it's important that these teachers also have information about their impact on student learning, and that this information contributes to their performance evaluation. In non-tested grades and subjects, states and districts can use such tools as nationally recognized assessments or district-wide, end-of-course exams and performance tasks. While these may be less statistically complex than the value-added measures used for tested grades, their focus on growth will ensure that

Only 42% of teachers agree that current evaluation allows accurate assessment of performance.⁵



they are appropriate measures of teacher impact on student learning.

Just as we wouldn't assign a semester grade for a student based on one assignment, we don't want one measure of student learning to carry the weight of the teacher's entire impact on student growth. For this reason, most new evaluation systems will require at least two measures of student learning as part of the evaluation. If a value-added calculation is one measure, the other could be a national, state or district-level assessment or performance task. Teachers in non-tested sub-

THE STUDENT PERSPECTIVE



The TRIPOD survey, administered by Cambridge Education, is an instrument that asks students to give feedback on specific aspects of a teacher's practice. This student perspective can provide teachers with invaluable information about how to improve their use of class time, their pedagogical practices, and their interactions with students. Surveys are customized for different age groups, but they all ask students to agree or disagree with detailed statements about their teachers' practices. Sample statements from the survey include:

"Our class stays busy and doesn't waste time."

"My teacher doesn't let people give up when the work gets hard."

"Students in this class treat the teacher with respect."

The Measures of Effective Teaching (MET) Project examines the correlations between the TRIPOD survey and teachers' impact on student learning. To address potential concerns that students who respond favorably on the survey may do so because they are high-achieving in the teacher's class, the researchers sought responses from one class of students taught by a teacher, but looked at the achievement gains of another group of students taught by the same teacher. The findings so far indicate that student perceptions are related to achievement gains. In other words, students can identify effective — and less effective — teaching when they experience it.⁶

QUALITY EVALUATION MATTERS

The TAP System for Teacher and Student Advancement works with schools across the country to implement comprehensive performance evaluation systems. In these systems, meaningful, carefully executed evaluation is a key tool used to lift teacher performance. Teachers are observed four to six times throughout the school year by multiple, trained evaluators. All observations are preceded by a meeting between the teacher and the evaluator, and are followed by a post-observation conference. During this conference, teachers receive detailed, actionable feedback followed by intensive, targeted professional development support.

Those conducting the observations include principals or other administrators, TAP master teachers, and TAP mentor teachers. All observers are highly trained and use an empirically validated rubric that measures performance on 19 indicators of effective instructional practice. TAP also monitors the raters' reliability, the validity of different evaluation elements, and ratings accuracy to ensure that all teachers are treated fairly and receive consistent ratings and feedback.⁷

The TAP evaluation process includes a measure of student learning, and a recent analysis found a strong correlation between TAP teachers' scores on observations and their impact on student learning.⁸ This suggests that these two measures, while assessing different aspects of performance, are, indeed, complementary.

TAP began more than 10 years ago and currently works with more than 10,000 teachers and 100,000 students across the country.⁹ Recent research shows that teachers working in TAP systems improve the quality of their instruction over time. TAP shines as an example of high-quality teacher evaluation and support that is improving teacher performance for the benefit of students.

“Receiving useful evaluations allows me to concentrate on my strengths, and to bring up my weaknesses, which benefits me and my students. After thinking that I was a good teacher, TAP showed me there were so many other ways that I could improve.”¹⁰

— Teacher, Knoxville, Tenn.

ject areas will also have multiple measures of student learning reflected in their evaluations.

In the coming years, some states and districts will move toward robust assessments in a wider range of subject areas. But in the interim, we can't ignore the impact scores of teachers have on students. New evaluation systems will find a way to measure the impact on student learning for all teachers, even if that approach does not look identical across all grades and subject areas.

A RICHER PICTURE

Classroom observations and measures of student learning are the most critical components of a good evaluation system. Most new evaluation systems will also include measures beyond these two in order to paint a richer picture of a teacher's performance, and to provide opportunities to highlight a teacher's strengths and areas for growth. For example, some recently implemented evaluation systems base a portion of the evaluation on teachers' contributions to the school community,



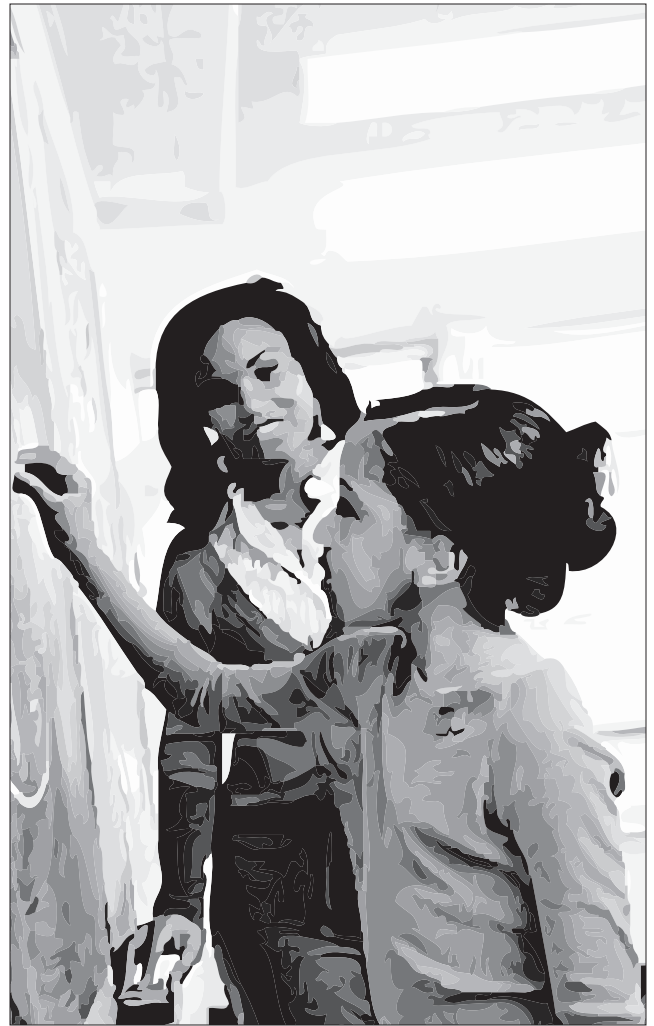
acknowledging the myriad responsibilities assumed by most teachers that exceed their classroom responsibilities. Some also include a schoolwide value-added measure to encourage teachers to see student improvement as a collaborative effort rather than an individual one. Other districts are looking into parent and student surveys, which can provide an intriguing perspective from which to measure a teacher's impact (see sidebar on page 6).

These additional measures are important in three ways. First, they provide even more context for a teacher's performance, improving the chances that a teacher's evaluation is a fair and rich representation of her ability. Second, they offer a lens that might be especially helpful for those grades and subjects that don't yet have clear tools to measure a teacher's impact on student learning. Third, and perhaps most significantly, many of these measures, such as student or parent surveys, can provide teachers with immediate, useful feedback about their use of time, pedagogical practices, classroom management, and interactions with students and parents.

LOOKING AHEAD

Better teacher evaluations will not solve all the challenges confronting education today. There are big and real concerns about how to build a more selective pipeline into the profession, and how to create conditions that keep great teachers in the schools and classrooms that most need them, among other issues. We should not pretend that improving evaluation systems will magically fix everything else. However, given the limitations created by current, broken systems of performance evaluation, fixing them is the most important step toward enabling us to address other problems strategically.

Melissa, the teacher highlighted at the beginning of this report, and all teachers like her deserve more than our current evaluation systems offer. They deserve a system centered on helping them to become even better teachers. They deserve a system that sets clear standards and expectations, and that gives them credit for the tremendous learning gains they make with their students. And they deserve a system that does not treat them as interchangeable parts but as professionals. Most important, as people who have dedicated their careers to helping students learn, they deserve to know that they are working in a system that is driven by what's best for students. And what's best for students is that teachers perform at the top of their game. Better evaluation systems are the best tool we can start with to help teachers get and stay there.



NOTES

- 1 Weisberg, D.; Sexton, S.; Mulhern, J.; and Keeling, D., "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness," (New York: The New Teacher Project, 2009).
- 2 Weisberg, Sexton, Mulhern and Keeling.
- 3 Coggshall, J., and Ott, A., "Retaining Teacher Talent: Convergence and Contradictions in Teachers' Perceptions of Policy Reform Ideas" (Naperville, Ill.: Learning Point Associates; and New York: Public Agenda, 2011).
- 4 "Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project" (Seattle, Wash.: The Bill & Melinda Gates Foundation, 2010).
- 5 Weisberg, Sexton, Mulhern and Keeling.
- 6 "Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project" (Seattle, Wash.: The Bill & Melinda Gates Foundation, 2010).
- 7 Jerald, Craig and Kristan Van Hook. "More than Measurement: The TAP System's Lessons Learned for Designing Better Evaluation Systems" (Santa Monica, Calif.: National Institute for Excellence in Teaching, 2011).
- 8 Daley, G. and Kim, L. "A Teacher Evaluation System That Works" (Santa Monica, Calif.: National Institute for Excellence in Teaching, 2010).
- 9 "TAP Research Summary," (Santa Monica, Calif.: National Institute for Excellence in Teaching, 2011).
- 10 "Voices from the Field: Teachers Describe Their Experience With a Bold System of Reform," (Santa Monica, Calif.: The National Institute for Excellence in Teaching, 2010)

ABOUT THE EDUCATION TRUST

The Education Trust promotes high academic achievement for all students at all levels — pre-kindergarten through college. We work alongside parents, educators, and community and business leaders across the country in transforming schools and colleges into institutions that serve all students well. Lessons learned in these efforts, together with unflinching data analyses, shape our state and national policy agendas. Our goal is to close the gaps in opportunity and achievement that consign far too many young people — especially those who are black, Latino, American Indian, or from low-income families — to lives on the margins of the American mainstream.

TheJoyceFoundation

The Education Trust is grateful to The Joyce Foundation for generously supporting our work in advancing teacher quality.



The Education Trust

1250 H STREET, NW, SUITE 700, WASHINGTON, D.C. 20005
P 202-293-1217 F 202-293-2605 WWW.EDTRUST.ORG