The Power of Performance Pay

Smarter teacher retention and
accelerated student achievement in Dallas

TEACHER EVALUATION reform dominated education policy throughout the 2010s, when new performance-based ratings were mandated in 44 states and Washington, D.C. Though high-stakes evaluation has since receded from the headlines, improving teacher quality remains a critical strategy to boost student outcomes and respond to new challenges, such as pandemic learning loss. So, it's worth taking a close look at the evidence on how performance-based evaluations can affect teacher quality and student achievement. What does the research show, and what can we learn by looking at districts that made the biggest changes?

One oft-cited 2021 study finds that high-stakes evaluations did not change teachers' paychecks and were, for the most part, a dud. Joshua Bleiberg and co-authors looked across the United States and found negligible effects from new evaluations that included multiple measures of teacher performance, including student test scores. Just like the perfunctory evaluations they replaced, many new systems rated less than 1 percent of teachers "unsatisfactory." In most states and districts, these systems also were disconnected from pay scales, which maintained traditional step-and-lane schedules that base teacher salaries on experience and education. Despite federal funding for incentives, evaluation reforms were too weak on their own to inform or induce meaningful changes in the quality of states' teacher workforces.

But that wasn't the case everywhere. Several large, urban districts implemented sweeping changes that linked performance-based evaluations with new, merit-based pay schedules. In Washington, D.C., for example, the IMPACT system rated teachers based on a variety of outcomes, including student test scores and professional observations, and triggered boosts in pay, targeted supports, or dismissal notices for educators at the ends of the spectrum. A long-running study by Thomas Dee and James Wyckoff found substantial improvement in teacher quality after IMPACT began in 2009, with greater retention of high performers and quick exits or improvements among teachers with lower performance rankings (see "A Lasting Impact," *research,* Fall 2017). Student achievement accelerated, particularly in math.

Over the past several years, we have investigated an even more comprehensive effort in Texas that, to date, has received far less attention. Starting in 2013, the Dallas Independent School District completely replaced its traditional pay scales for principals and teachers with an evaluation and compensation system based on multiple measures of effectiveness, including student achievement and student survey responses. The district also established new, robust definitions of

By ERIC A. HANUSHEK, MINH NGUYEN, BEN OST, and STEVEN G. RIVKIN

Natasha Boone, a 3rd-grade reading teacher, high fives a student at Titche Elementary School in Dallas in 2019. Boone was one of 400 teachers who participated in Dallas ISD's successful ACE program to turn around the district's lowest-performing schools.

educator excellence, performance-based reviews for school principals, and cash incentives to encourage highly rated teachers to move to low-performing schools.

We conducted multiple analyses to track the impact of these efforts. The results show the district's reforms had a large and durable positive impact on teacher quality and student learning.

In the four years after Dallas adopted new performance-based teacher evaluation and compensation systems, student performance on standardized tests improved by 16 percent of a standard deviation in math and 6 percent in reading, while scores for a comparison group of similar Texas schools remained flat. Teacher turnover in the wake of these reforms was concentrated among lower-rated teachers. And a program that offered sizable financial incentives to reassign top-rated teachers to struggling elementary campuses immediately improved teacher quality and student achievement and had dramatic, lasting, positive effects on student learning through middle school.

Evaluation and Pay Reform in Dallas

A large, urban school district in north central Texas, Dallas ISD enrolls roughly 139,000 students in 240 schools. Some 72 percent of students are Hispanic, about 20 percent are Black, and about 6 percent are white. Approximately 90 percent of students are eligible for free or reduced-price school lunch, and the four-year graduation rate is around 80 percent, which is below the statewide average.

Local efforts to change educator evaluation and compensation began in earnest in 2011, after new state rules empowered Texas districts to develop their own ways of rating teacher performance. In Dallas, the district board of trustees adopted a pay-for-performance compensation system proposed and developed by then-Superintendent Mike Miles. Over about three years, the district established a new multiple-measures evaluation system based on classroom observations, growth in student test scores when available, and student surveys.

The evaluations, adopted in 2015 as part of the Teacher Excellence Initiative, or TEI, are based on detailed rubrics defining excellence and on aligned professional development for teachers and principals. A parallel reform for principals, the Principal Excellence Initiative, uses a similar method to assess and categorize principals by performance, including their use of the rich information created by TEI evaluations to help teachers improve. Pay for teachers and principals is based on their evaluation scores averaged over two years. In combination, these structures aim to support educator growth, to strengthen incentives to improve instruction and leadership practices, and to attract and retain strong teachers and school leaders in Dallas ISD.

Teacher evaluations include 10 classroom observations (some unannounced) each year by the same observer, evidence of student progress toward established learning objectives, test-based measures of achievement growth relative

In the four years after Dallas adopted new performance-based teacher evaluation and compensation systems, student performance on standardized tests improved by 16 percent of a standard deviation in math and 6 percent in reading.

to comparable students, and schoolwide achievement. The district also surveys students in grades 3 through 12 each spring and incorporates responses into eligible teachers' performance ratings.

Each year, teachers receive an evaluation score that is used to assign them to a performance rating: unsatisfactory, progressing I and II; proficient I, II, and III; and exemplary. Performance-based salaries in the first year of TEI ranged from $45,000 to $90,000, with the largest share of teachers paid $54,000 at the proficient I level. The system maintained fixed proportions of teachers in each performance category; for example, the exemplary category is targeted for teachers in the top 2 percent by evaluation score, while the unsatisfactory rating is targeted for teachers in the bottom 3 percent. A teacher cannot move up or down more than one effectiveness level per year, and a teacher's salary can only be adjusted downward after they score at a lower level for three consecutive years.

In 2016, the district built on this work through the Accelerating Campus Excellence program, or ACE, which offers up to $10,000 in additional pay for the highest-rated teachers as an incentive to work in the lowest-performing schools. ACE teachers also are required to work longer hours, use data-driven instruction, and pass ongoing, rigorous screenings to remain in the program, which resulted in the rapid and voluntary reassignment of most ACE educators in a single year.

We assess the impacts of the Dallas ISD reforms by looking at overall student performance data on state tests in math and reading during a four-year period from 2015 to 2019. We conduct a second analysis focused on schools included in the ACE program. We also look at rates of differential teacher retention based on performance ratings and estimate the degree to which a more effective teaching force contributed to changes in student achievement.

Our analyses are based on student enrollment and demographic data; teacher and principal data such as role, experience,

salary, education, class size, grade, population served, and subject taught; and student performance on annual statewide tests in grades 3 through 8. Unique student and educator identifiers enable us to follow students and teachers across districts and schools as long as they remain in a Texas public school. We also create a comparison group from elementary and middle schools in high-poverty Texas school districts that did not adopt evaluation and compensation reforms.

Impacts on Student Achievement

After Dallas ISD implemented multiple-measure teacher evaluations and performance-based compensation, students did significantly better on statewide math and reading exams. By 2019, student achievement in math improved by 16 percent of a standard deviation; reading achievement improved by 6 percent of a standard deviation (see Figure 1).

These results come from looking at student performance over time compared to a synthetic comparison group of schools drawn from other high-poverty Texas districts. In tracking the impacts of evaluation and compensation reform over time, we find no difference between Dallas ISD and the comparison group until 2016, the second year of the teacher evaluation and compensation reforms. After that, Dallas scores steadily rise through 2019 (the last year before the Covid-19 pandemic). The initial lag in impact is not surprising given the design of the reforms, which were built on incentivizing, supporting, and rewarding high performance in the classroom. Since evaluations began in 2015, any resulting difference in overall teacher quality would not begin until 2016.

Was differential retention of high- and low-performing teachers the driving force behind these improvements? The Dallas reforms involved simultaneous changes in the strength of incentives, information available for mentoring and professional development, and myriad aspects of school operations and educator composition, complicating efforts to disentangle the contributions of each. That said, if the much closer alignment between effectiveness and salary altered the composition of entrants to and exits from Dallas ISD, educator composition could have been an important channel through which the reforms improved student outcomes in the district. A first-order issue, therefore, is understanding the impact of the reforms on educator selection.

Increases in Student Achievement After Evaluation and Pay Reform *(Figure 1)*

After Dallas ISD enacted new teacher evaluations and performance-based pay, student scores on statewide math tests improved by 20 percent of a standard deviation over four years. Scores at a comparable group of high-poverty Texas schools that did not reform evaluation or compensation were flat. The trends in reading were similar.
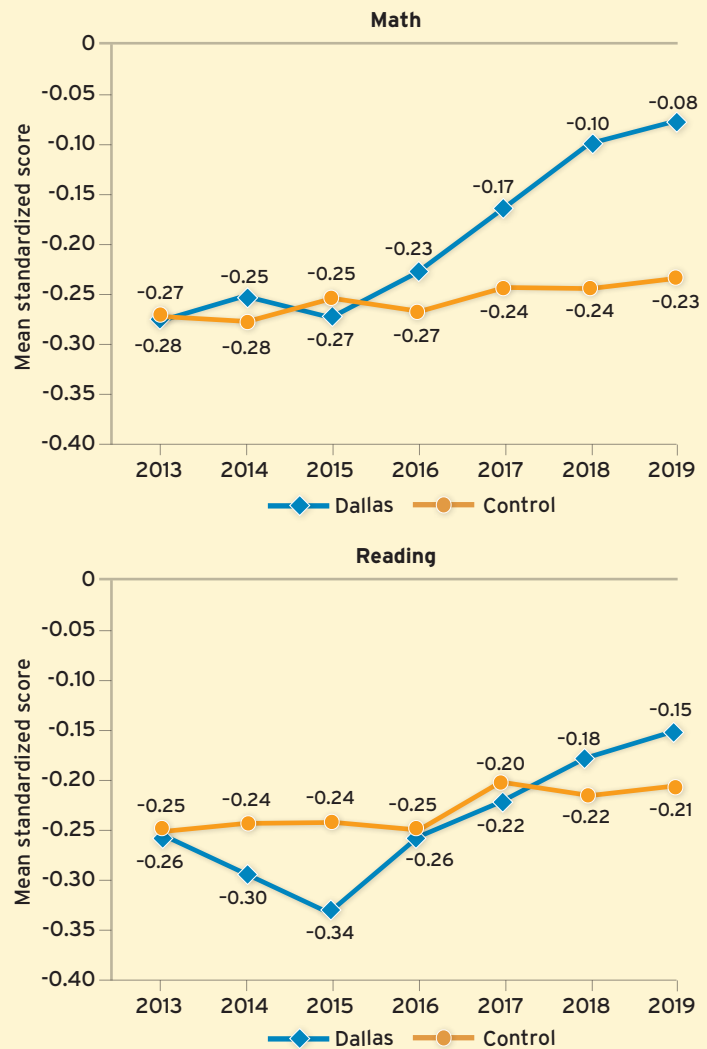
NOTE: Changes in standardized test scores in math and reading in Dallas ISD elementary and middle schools compared to a synthetic control district created by the authors. Test scores are standardized with a mean of zero and variance equal to one to depict student performance relative to the overall state performance distributions. Control is composed of schools from other large, high-poverty Texas districts.

SOURCE: Authors' calculations

The Role of Teacher Turnover

We focus on teacher departures from Dallas ISD to understand the effects of evaluation and compensation reform on the district's workforce rather than looking at new arrivals for a practical reason: no other district uses comparable measures of effectiveness. Even estimates of teacher value-added to student test scores, which is a common measure, are available only for the small fraction of entrants who previously taught in a tested grade in another district. No effectiveness measures are available for new entrants to teaching.

The rate of teacher turnover rose sharply after 2012, when the district's controversial reform efforts were highly publicized but still in the development stage. This increase produced major shifts in the shares of teachers with minimal experience. The share of novice teachers with no prior experience quadrupled within three years, from 3 percent in 2012 to 13 percent in 2015. The share of early-career teachers with zero to two years of experience grew sharply from 12 percent in 2012 to 32 percent in 2016 and then declined modestly until 2019. Because new teachers' effectiveness improves rapidly in their first few years in the classroom, this influx of teachers to Dallas ISD who had little or no prior experience likely had a negative effect on achievement that temporarily dampened achievement growth relative to the synthetic control.

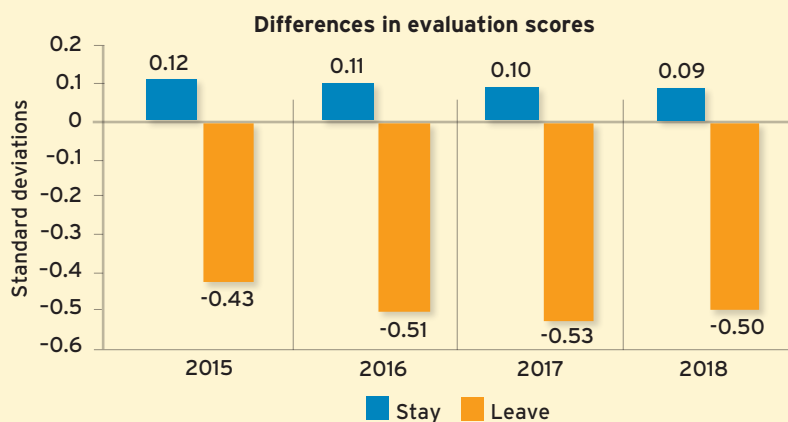However, the implications of higher turnover depend on

A comparison of average evaluation scores for teachers who left Dallas ISD with those who stayed on the job shows pronounced negative selection out of the district. The average evaluation score for teachers who remain in Dallas ISD exceeds those who leave by more than 50 percent of a standard deviation starting in 2016.

whether exiting teachers are above or below average. Although a low rating didn't trigger dismissal, it did come with a potential negative impact on pay and could have led poor performers to leave on their own accord. We turn our attention to 2015, when TEI took effect, and the years immediately after and then compare the average evaluation scores for teachers who left the district and those who stayed on the job. This comparison reveals pronounced negative selection out of the district (see Figure 2). The average evaluation score for teachers who remained in Dallas ISD exceeded those who left by more than 50 percent of a standard deviation starting in 2016.

Whether the departure of less effective educators translates into better instruction depends on the quality of their replacements. The absence of a measure of effectiveness for teachers prior to their entry into Dallas ISD precludes the direct estimation of the change in teacher effectiveness; however, we perform a separate analysis to estimate the overall contribution of changes in the composition of the teacher workforce to the district's student achievement gains. Composition of the teaching force is estimated to contribute more than half of the impact on student learning, in combination with other factors including strengthened performance incentives, enhanced support based on

Highly Rated Teachers Earn More and Are More Likely to Stay on the Job *(Figure 2)*

Under Dallas ISD's performance-based compensation system, the evaluation scores of teachers who are retained and return to the classroom the following year are more than 50 percent of a standard deviation higher than scores for teachers who exit the district.

Differences in evaluation scores

Year	Stay	Leave
2015	0.12	-0.43
2016	0.11	-0.51
2017	0.10	-0.53
2018	0.09	-0.50

Standard deviations

SOURCE: Authors' calculations

detailed classroom observations and evaluation data, and more effective instructional and school leadership.

Attracting Effective Educators to Hard-to-Staff Schools

In 2016, Dallas ISD built on its innovations in measuring and rewarding teacher performance to address the challenge of attracting and retaining effective teachers in hard-to-staff, chronically low-performing schools. The path-breaking ACE program focused on selectively retaining and recruiting very high-performing teachers and used large pay increases to reshape instructional staff at schools serving disadvantaged students. It was launched at the district's four lowest-scoring elementary schools in 2016 and expanded to nine schools in 2018.

At the program's outset, less than 20 percent of existing staff in ACE schools met ACE performance standards and were retained. The remaining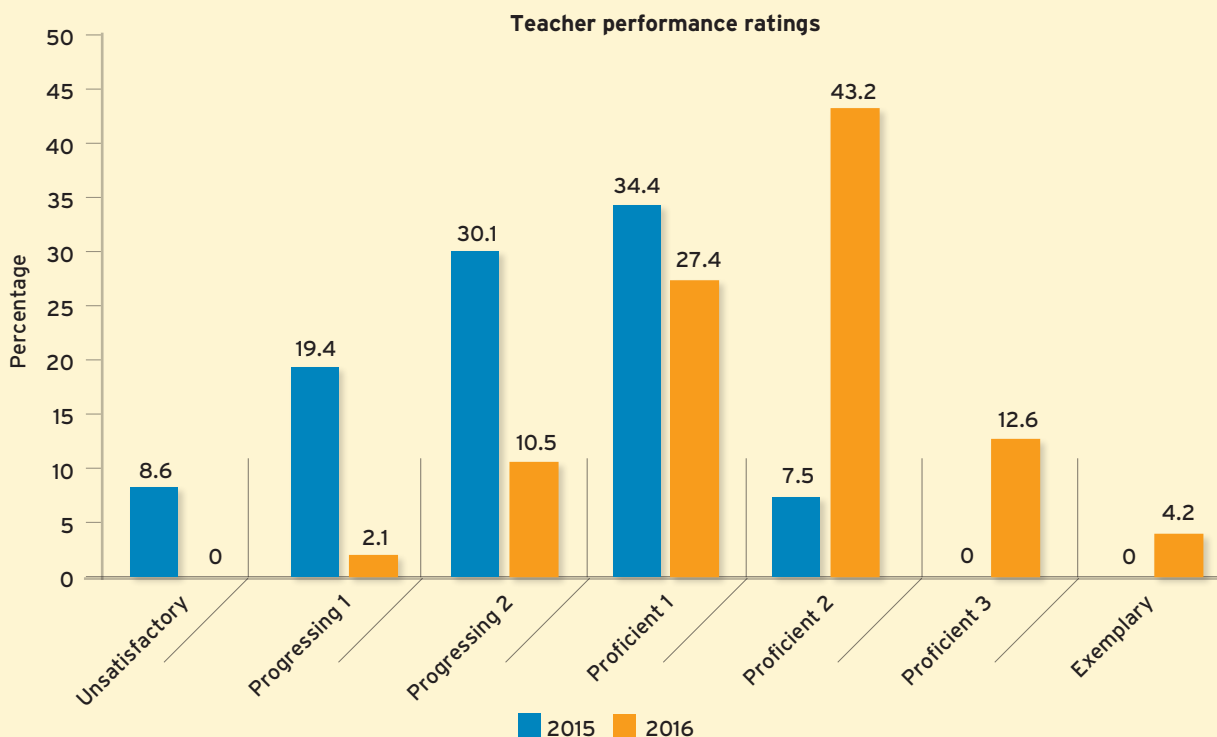 positions were filled by highly rated teachers who transferred from other schools. Teachers who applied and were selected to work at ACE campuses received signing bonuses of $2,000 and annual stipends between $6,000 and $10,000 depending on their position and effectiveness rating from the previous year. Principals, counselors, and instructional coaches received stipends that ranged from $6,000 to $13,000 annually.

We look at the ratings of teachers in ACE schools before and after the program's start in 2016, and the shift is transformational. Before ACE, the vast majority of teachers were rated in the bottom three categories of performance; after ACE, more than half were rated in the top three performance categories (see Figure 3).

Changing school staffs is not the only focus of the program. Under ACE, educators use data-driven instructional practices and are subject to rigorous performance screenings to retain their roles. Students at ACE schools are provided with three meals a day, afterschool enrichment, and other developmental

Signing Bonuses and Stipends Shift Teacher Quality in ACE Schools *(Figure 3)*

Before the ACE program, more than half of teachers in participating schools were rated below proficient. After the program offered signing bonuses and performance-based stipends of up to $10,000 to encourage highly-rated teachers to transfer, more than half of teachers in participating schools were in the top three performance categories.

Teacher performance ratings

Rating	2015	2016
Unsatisfactory	8.6	0
Progressing 1	19.4	2.1
Progressing 2	30.1	10.5
Proficient 1	34.4	27.4
Proficient 2	7.5	43.2
Proficient 3	0	12.6
Exemplary	0	4.2

supports. These interventions and teacher stipends remain in place until student achievement improves and the school no longer qualifies for the program.

To assess the impact of ACE on student learning, we compare scores on standardized reading and math tests at ACE schools with a similar group of Dallas ISD elementary schools with 2014 test scores in the lowest 15 percent of the district. We focus our analysis on three elementary schools in the first wave of the ACE program, from 2016 to 2018. (Scoring problems precluded looking at the fourth ACE school.)

ACE schools show an immediate, large increase in achievement while scores at comparison schools are flat (see Figure 4). Scores at ACE schools increase by almost 50 percent of a standard deviation in math and 25 percent of a standard deviation in reading in the first year and continued to improve in years two and three, when ACE stipends and supports remained in place. Performance in comparison schools improve in those years as well, in line with overall district improvement, but the increase is less steep than for the ACE schools.

In 2019, student achievement at all but one ACE school had improved such that the schools were removed from the program and teacher stipends and additional instructional time ended. After that occurred, teacher quality and student achievement experienced sharp declines: more than 40 percent of teachers rated proficient 1 or higher left the ACE schools and average test scores fell by 23 percent of a standard deviation in math and 17 percent of a standard deviation in reading. Achievement at comparison schools was largely unchanged.

Importantly, students who attended ACE elementary schools during that time experienced lasting positive effects seen in subsequent middle school performance. Students who were in 3rd grade when the program began and received three years of ACE supports score 39 percent of a standard deviation higher in math and 23 percent of a standard deviation higher in reading in 6th grade than similar students in comparison schools (see Figure 5). The prior score gains are not just the result of "teaching to the test" but represent true learning gains.

Implications

The Dallas reforms prove what's possible when teacher evaluation and compensation reform are part of a comprehensive reset of districtwide personnel policies and practices. The district virtually eliminated the dependence of salary on experience and postgraduate degrees,

Student Achievement Increases with Support and Teacher Stipends *(Figure 4)*

Students' test scores improved by more than 40 percent of a standard deviation in math and 20 percent of a standard deviation in reading in the first year of ACE, which used financial incentives to replace low-rated staff with highly rated teachers and offered robust wrap-around services and supports. In 2019, when ACE schools improved and no longer qualified for the intervention, student achievement fell. Achievement was relatively flat for a control group of similarly lower-achieving Dallas ISD schools that did not participate in ACE.
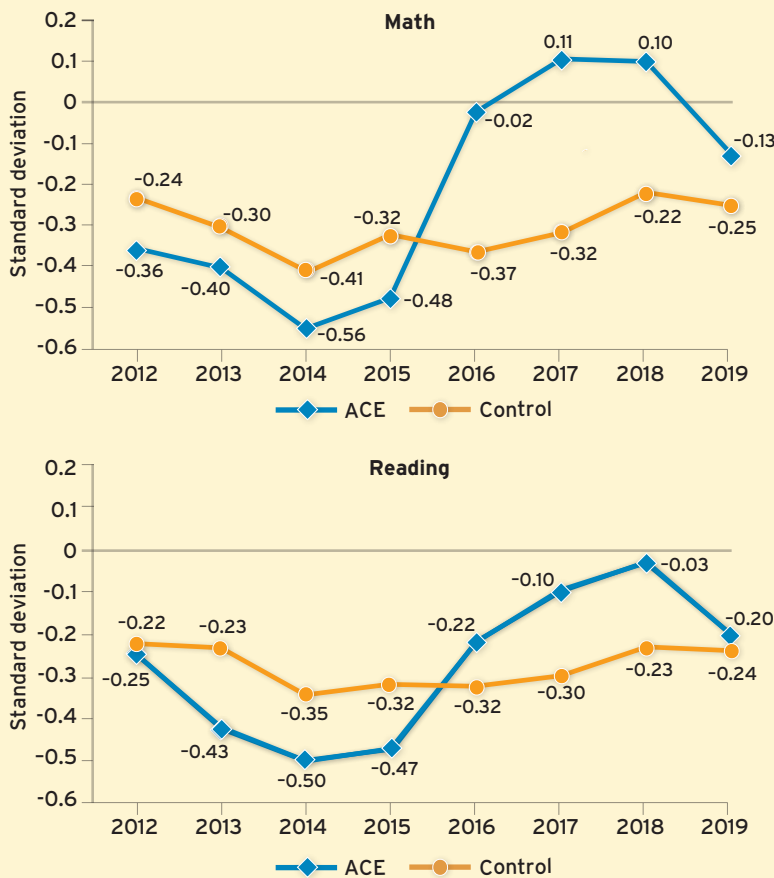
NOTE: Data for ACE schools does not include Umphrey Lee Elementary School.
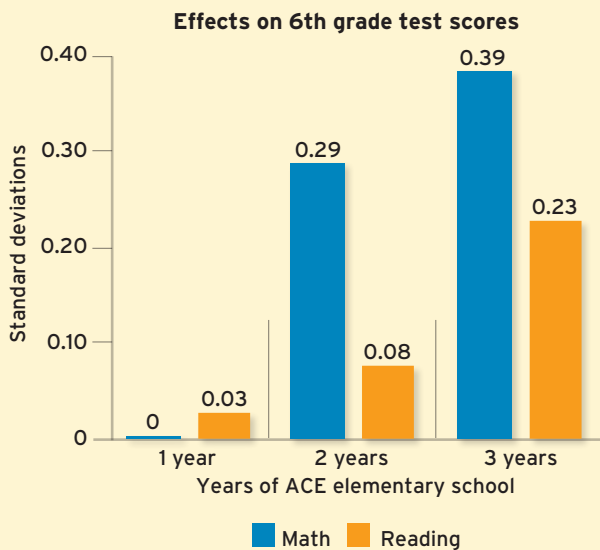
SOURCE: Authors' calculations

radically altering the traditional systems of evaluation and pay found throughout the United States. As a result, both teacher quality and student achievement improved.

The ACE program shows how reforms can be targeted to address the needs of chronically low-performing schools. The information produced by Dallas ISD's evaluation and compensation reforms provided the basis for effectiveness-adjusted hiring and pay in hard-to-staff schools. Teachers respond to incentives. Our analysis shows that the ACE program remade school staffs virtually overnight and boosted student learning, such that the poorest-performing elementary schools moved close to the district average in just two years. Though that success ultimately resulted in the removal

Lasting Impacts from Assigning High-Rated Teachers to Low-Performing Schools *(Figure 5)*

Financial incentives that bring highly rated teachers to low-performing elementary schools have positive effects on achievement in 6th grade, with the biggest impacts among students who are in 3rd grade at the start of the program. Students who experience three years of ACE score 39 percent of a standard deviation higher in math and 23 percent of a standard deviation higher in reading compared to similar students who attended low-performing schools not in the ACE program.

Effects on 6th grade test scores

SOURCE: Authors' calculations

Before ACE, the vast majority of teachers were rated in the bottom three categories of performance. After ACE, more than half were rated in the top three performance categories.

of schools from the program, students who experienced the ACE reforms continued to benefit into middle school.

While such sweeping changes may appear blunt from a distance, a close look at the Dallas reforms shows they were carefully planned to guard against evaluation inflation, the arbitrary treatment of teachers, and strategic responses such as teaching to the test. Aligning the relationship between educator effectiveness and pay dramatically strengthened performance incentives, while the development of a multiple-measure evaluation system that includes evidence of student learning, supervisor observations, and student-survey feedback recognized the pitfalls of a singular reliance on either test scores or subjective evaluations by supervisors. Importantly, focusing on teachers' value-added rather than absolute performance measures like passing rates or achievement benchmarks made it clear that the district sought to account for factors outside of educators' control. As a result, these systems survived controversy and contributed to substantial gains in teacher quality and student learning.

Indeed, this experiment in improved personnel policies continues and has expanded. The State of Texas introduced a grant program designed to induce other districts to follow Dallas's lead, and some 400 districts have begun such a transformation. And in 2023, the state took over Houston ISD and appointed Mike Miles—the architect of the Dallas system—as superintendent. The largest district in Texas is now undergoing similar evidence-based changes in personnel policies.

Eric A. Hanushek is the Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University; Minh Nguyen is an assistant professor of economics at Ball State University; Ben Ost and Steven G. Rivkin are professors of economics at University of Illinois Chicago. This article is based on two working papers published by the National Bureau of Economic Research: "The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement" by Hanushek and co-authors and "Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools" by Andrew J. Morgan, Hanushek, and co-authors.