

Efforts to Toughen Teacher Evaluations Show No Positive Impact on Students



More than a decade ago, policymakers made a multi-billion-dollar bet that strengthening teacher evaluation would lead to better teaching, which in turn would boost student achievement. But new research shows that, overall, those efforts failed: Nationally, teacher evaluation reforms over the past decade had no impact on student test scores or educational attainment.

The research is the latest indictment of a massive push between 2009 and 2017, spurred by federal incentives, philanthropic investments, and a nationwide drive for accountability in K-12 education, [to](#)

[implement high-stakes teacher evaluation systems](#)
nearly every state.

in

Prior to the reforms, [nearly all teachers received satisfactory ratings](#)

in their evaluations. So policymakers from both political parties introduced more-robust classroom observations and student-growth measures—including standardized test scores—into teachers' ratings, and then linked the performance ratings to personnel decisions and compensation.

“There was a tremendous amount of time and billions of dollars invested in putting these systems into place, and they didn't have the positive effects reformers were hoping for,” said Joshua Bleiberg, an author of the study and a postdoctoral research associate at the Annenberg Institute for School Reform at Brown University. “There's not a null effect in every place where teacher evaluation [reform] happened. ... [But] on average, [the effect on student achievement] is pretty close to zero.”

The evaluation reforms were largely unpopular among teachers and their unions, who argued that incorporating certain metrics, like student test scores, was unfair and would drive good educators out of the profession. Yet proponents—including the Obama administration—argued that tougher evaluations could identify, and potentially weed out, the weakest teachers while elevating the strongest ones.

“We think the goal of great teaching is to have students learn; and to have student learning be a piece of teacher evaluation, I think, actually gives the profession the respect it deserves,” said Arne Duncan, who served as President Obama's education secretary from 2009 to 2016, [in an EdWeek interview in](#)

2015

But teachers said the focus on student growth measures stripped away the emphasis on building relationships with students.

“It took away the overall focus on the kid and the overall focus on teaching,” said Erin Scholes, an innovation coordinator at a Connecticut middle school who has been in the classroom for 15 years. “I felt like [the reforms] hit the science of teaching rather than the art of teaching and tried to fit everyone in the same box.”

Researchers found no positive effects on student outcomes

A team of researchers from Brown and Michigan State Universities and the Universities of Connecticut and North Carolina at Chapel Hill analyzed the timing of states' adoption of the reforms alongside district-level student achievement data from 2009 to 2018 on standardized math and English/language arts test scores. They also analyzed the impact of the reforms on longer-term student outcomes, including high school graduation and college enrollment. The researchers controlled for the

adoption of other teacher accountability measures and reform efforts taking place around the same time, and found that their results remained unchanged.

They found no evidence that, on average, the reforms had even a small positive effect on student achievement or educational attainment.

The study's authors noted that the design and implementation of the reforms fell short of the recognized best practices for performance management systems. Under a program known as Race to the Top, the Obama administration offered states \$4.35 billion in competitive grants for enacting certain policy changes, including incorporating student achievement data in their evaluation systems. The government also used a waiver system that would allow states to receive some regulatory relief from stringent federal requirements if they implemented more accountability measures for teachers.

But in practice, implementation proved difficult in most places, with [most teachers still receiving](#)

[satisfactory ratings](#) under the new evaluation systems. Performance-based dismissals were still rare, and states that linked evaluation ratings to compensation often offered only small bonuses or set the bar so low that most teachers qualified.

Also, the reforms decreased job satisfaction among new teachers who felt like they had little autonomy to do their best work, the paper noted. And they added significant demands to administrators' already burdensome workload.

"It was really the worst of all worlds," said Michael Petrilli, the president of the Thomas B. Fordham Institute, a conservative education think tank that advocated for more teacher accountability. "It was just a big paperwork exercise. It led to a lot of anxiety and bad morale. Not only did it have no findings [of positive effects on student outcomes], it had real-world consequences that were almost entirely negative."

Tougher teacher-evaluation systems can work, Petrilli said—but there was no political will to act on the results at the time of the reforms. Teachers' unions resisted firing teachers who received poor results, and districts were unwilling or unable to pay great teachers more, he said.

Indeed, past research done in 2017 found that [principals continued to rate nearly all teachers as](#)

[effective](#), even though researchers found the principals would give harsher ratings in confidence with no stakes attached.

"We just don't have a system in the country that's well set up to push the rapid implementation of any education reform, including teacher evaluation," Bleiberg said. "You see a lot of superficial adoption—that's likely to lead to the null effects overall."

Evaluation reform has already changed course

States overhauled their teacher-evaluation systems quickly, and then many reversed course within just a few years. A [National Council on Teacher Quality analysis](#) found that

the number of states that required student-growth data in teacher evaluations went from 15 in 2009 to 43 in 2015—and then back down to 34 in 2019.

The changes were in part due to the increased flexibility states now have under the Every Student Succeeds Act, which stripped the U.S. secretary of education of the power to determine how states grade their teachers.

Also, other research into the outcomes of evaluation reform has produced similarly discouraging results. For example, a \$575 million effort, funded in part by the Bill & Melinda Gates Foundation, to implement new teacher-evaluation systems in three large school districts was found to have been largely ineffective in increasing student achievement.

Experts say the results show the difficulties of implementing any large-scale reform, but in particular a top-down model that was forced onto districts and adopted without much buy-in from those on the ground. And some say the evaluation reforms were done without considering other constraints on the profession.

“Yes, most of our teachers could be better at their jobs, but it’s not because they’re not trying hard enough,” said Jack Schneider, an associate professor of leadership in education at the University of Massachusetts Lowell. “It’s because they teach too much, they have too many students in their classrooms, they don’t have relevant and sustained professional development opportunities, they don’t have adequate support from school leaders who themselves are overburdened in schools. There’s a lot we could do if we wanted to strengthen the teaching profession, but most of these reforms didn’t really address the fundamental barriers that keep teachers from being their best professional selves.”

The reforms were also demoralizing for teachers, said Rebecca Garelli, a science education consultant who taught for 14 years and left the classroom partly because of the increased focus on student test scores.

“To tie those test scores to my evaluation was something I innately struggled with from the beginning,” she said. “It never made sense to me to take something so human and turn it into something so non-human.”

Even so, there are bright spots in teacher-evaluation reform, many say, most notably in Washington, D.C. The district’s teacher-evaluation system, known as IMPACT, ties student test scores to teachers’ job security and paychecks. Under the system, teachers who receive “ineffective” scores are subject to dismissal, and teachers who score “minimally effective” or “developing” could face dismissal if they don’t improve. “Highly effective” teachers, however, are eligible for financial rewards and professional opportunities.

Research has found that lower-performing teachers in the District of Columbia school system are more likely to voluntarily leave than their higher-performing counterparts. When they leave, they are

replaced by teachers with higher IMPACT scores, [and student achievement increases](#)

. And when they do stick around, their performance tends to improve.

Other states and districts used similar evaluation systems, but there were some key differences, the study's authors said.

The former D.C. school chancellor, Michelle Rhee, and the local teachers' union had a long, bitter dispute about the details of evaluation reform, but [eventually the two sides worked out an agreement](#)

, with both sides making concessions, Bleiberg said. (Even so, the teachers' union says the evaluation system has created a culture of fear in the district. And a recent study found that [the system is racially biased](#)

, with white teachers on average receiving higher scores than their Black and Hispanic peers.)

In many places, governors didn't work with teachers' unions before implementing evaluation reforms, Bleiberg said: "It was a reform that was all about teachers and didn't really end up getting them on board."

'We know it's possible' to achieve positive outcomes

Still, the results in Washington and in other cities show that high-stakes teacher-evaluation systems can work, said Kate Walsh, the president of the National Council on Teacher Quality, a Washington-based group in favor of measuring teacher effectiveness through objective data like test scores.

"We know it's possible for teacher evaluation [reform], when well-implemented, to achieve great outcomes," she said. "We know it's theoretically possible, and we know it's practically possible."

But there's little evidence to suggest a large number of school districts can meaningfully implement any sort of reform and get positive results, Walsh said, especially in a relatively short amount of time.

“I think people were serious about it for two years max—you’re not going to get good outcomes in a couple years,” she said. “You have to do it a while before you can reap the benefits.”

Also, teacher-evaluation systems cannot be changed in a vacuum, said Garrett Landry, the founder and CEO of Steady State Impact Strategies, a consulting firm working with school districts in Texas to reform the way they identify—and reward—effective teachers.

Teachers have to have the right conditions for success, he said, and improving teacher quality has to start with ensuring principal quality. Landry said districts should anchor their teacher-evaluation systems in growth and delineate clear targets for teachers to meet.

“We don’t really have time [to waste] in education. . . . If we don’t get [students] on track early, it’s really hard to catch them up,” he said. “We really need the best and brightest educators, and too many systems can’t tell me who the best educators are. Everybody looks the same on paper.”

There’s currently little political appetite to try again with teacher-evaluation reform, Bleiberg said. That’s in part due to the pandemic, [which has dampened teacher morale](#)

, but he also thinks policymakers will need to take time to generate more buy-in and address the fundamental challenges of implementation.

But Walsh said the issue will come up again, as part of the cyclical nature of school reform.

“It’s not acceptable to have an evaluation system where everyone gets the same rating,” she said. “Because we didn’t do it well [the last time] doesn’t mean it can’t be done well. We’ve just got to find a different way.”