

'An Expensive Experiment': Gates Teacher-Effectiveness Program Shows No Gains for Students

By **Madeline Will**

June 21, 2018

The Bill & Melinda Gates Foundation's multi-million-dollar, multi-year effort aimed at making teachers more effective largely fell short of its goal to increase student achievement—including among low-income and minority students, a new study found.

 [Back to Story](#)

This conclusion to an expensive chapter of teacher-evaluation reform shows the difficulty of making sweeping, lasting changes to teacher performance. The results also demonstrate the challenges of getting schools and teachers to embrace big changes, especially when state and local policies are in flux.

The evaluation of the program, **released today**, was conducted by the RAND Corporation with the American Institutes for Research and was funded by the Gates Foundation.

Under its intensive partnerships for effective teaching program, the Gates Foundation gave grants to three large school districts—Memphis, Tenn. (which merged with Shelby County during the course of the initiative); Pittsburgh; and Hillsborough County, Fla.—and to one charter school consortium in California starting in the 2009-10 school year. The foundation poured \$212 million into these partnerships over about six years, and the districts put up matching funds. The total cost of the initiative was \$575 million.

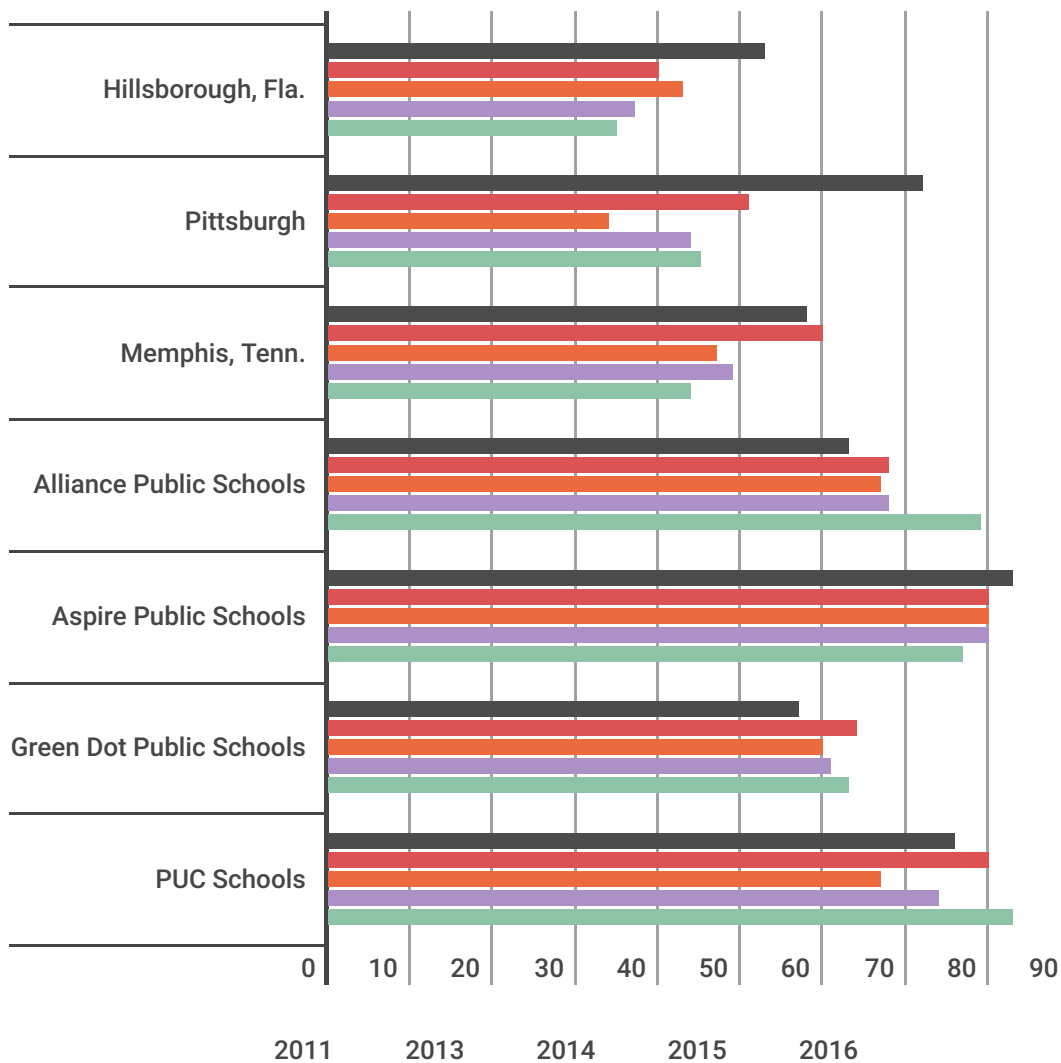
The school sites agreed to design new teacher-evaluation systems that incorporated classroom-observation rubrics and a measure of growth in student achievement. They also agreed to offer individualized professional development based on teachers' evaluation results, and to revamp recruitment, hiring, and placement. Schools also implemented new career pathways for effective teachers and awarded teachers with bonuses for good performance.

"The initiative itself tried to pull a bunch of levers to have a big impact on student performance," said Brian Stecher, a RAND researcher and the lead author of the report. "The sites did in fact modify all of these levers, some more than others, but in the end, there were no big payoffs in terms of improved graduation [rates] or achievement of students in general, and low-income and minority students in particular."

Do Teachers Think Students Will Benefit From the Evaluation Reform?

Researchers tracked participating teachers' agreement that in the long run, students will benefit from the teacher-evaluation system. The proportion of teachers who thought it would help students declined since 2011 in the three public school districts—most dramatically in Pittsburgh.

Enthusiasm is higher among teachers in the charter schools.



*Researchers did not survey teachers in 2012.

SOURCE: RAND Corporation and American Institutes for Research

EDUCATION WEEK

By the end of the 2014-15 school year, the study found, student outcomes were not significantly better than outcomes in similar school sites that did not participate in the initiative. Researchers also found no evidence that low-income and minority students had greater access to effective teachers than their white, more-affluent peers, which had been another stated goal by the Gates Foundation. (Researchers also collected student outcome data for the 2015-16 and 2016-17 school years, and will update the conclusions this fall or next spring.)

A caveat to these results is that while the initiative was taking place, high-stakes teacher-evaluation measures were also being enacted across the country. This made it difficult to tease out the results of the Gates-led teacher-evaluation systems, compared to what was being implemented elsewhere. The research looked at the extent to which the Gates partnerships improved student outcomes over and above the statewide reforms.

Still, at the end of the research period, very few teachers in participating districts were classified as ineffective, which researchers believe is in part due to an unwillingness among school leaders to give

harsh ratings based on classroom observations. Also, sites did not ultimately retain more effective teachers, although researchers did find declines in the retention of ineffective teachers.

"We believe that this work, which originated in ideas that came from the field, led to critical conversations and drove change and partnerships across the country," said Allan Golston, the president of the U.S. program at the Gates Foundation, in a statement. "We have taken these lessons to heart, and they are reflected in the work that we're doing moving forward."

Last October, the Gates Foundation had **announced a major shift in its investment strategy for education**, pivoting away from teacher-evaluation efforts entirely. The foundation plans to pump \$1.7 billion into K-12 education, with a focus on improved curricula that match state standards for learning and helping networks of middle and high schools scale up best practices.

"We'll no longer directly invest in teacher evaluation, but we'll continue to gather data on the impact of these systems and encourage the use of all of those tools that help teachers improve their practice," said Bill Gates in his speech announcing the new investment. **Preliminary results** of the intensive teaching partnership had indicated that the work was not translating into widespread achievement gains for students.

(*Education Week* receives financial support from the Gates Foundation for coverage of continuous improvement strategies in education, and has received grant funding in the past for coverage of college- and career-ready standards implementation. *Education Week* retains sole editorial control of its content.)

'Pushback and Disharmony'

Before this latest pivot, Gates had devoted **at least \$700 million** to its teacher-quality agenda, including a massive, three-year study of **how to measure effective teaching** that concluded in 2013. That prior study—the results of which were incorporated into this more-recent partnership work—demonstrated that great teaching can be identified through classroom observations, student surveys, and student test scores.

However, the singular focus on teacher effectiveness in the partnership work might be one reason student achievement didn't improve, Stecher said.

"This suggests that focusing on [teacher effectiveness] alone is not likely to be the potent sort of intervention that really moves the needle on student outcomes," he said, adding that maybe factors like early-childhood education, family support, and child nutrition also need to be addressed to make a significant impact on student performance.

For that reason, many educators weren't entirely surprised by the results, said Ted Dwyer, who is the chief of data, research, accountability, and assessment at Pittsburgh schools.

"A lot of people in districts felt like there was a disconnect, and [the initiative] created an enormous amount of focus on the adults in the system, rather than what really matters in the system, which is our kids," said Dwyer, who was the manager of evaluations at the Hillsborough district when the Gates work first started.

Still, years of research show that teacher effectiveness is important for student growth, said Daniel Goldhaber, the director of the Center for Education Data and Research at the University of Washington, who has studied issues of teacher performance for more than a decade and whose center receives Gates funding. (Goldhaber is also employed by AIR, but was not involved in this research.)

"These findings don't undermine any of the papers that this [initiative] was built on," he said. "It undermines the notion that we have the political will to do this."

Indeed, the RAND study found that while all sites initially had approval from most involved parties to adapt their teacher-evaluation systems, teachers' unions began to object a few years into the process.

Teachers' Perceptions of Evaluation Systems

In spring 2016, researchers asked teachers whether they thought the consequences tied to evaluation results are reasonable, fair, and appropriate. Consequences can include possible termination or increased compensation.

Teachers in charter schools were much more likely to believe the high-stakes consequences were fair.

SOURCE: RAND Corporation and the American Institutes for Research

EDUCATION WEEK

"When the results started being used to give cash rewards or to identify teachers for required planning and ultimately, perhaps, termination, the teacher organizations reacted defensively," Stecher said. "[Districts] had to suffer through a lot of pushback and disharmony."

And that ill will might have influenced evaluation scores, the study suggests. Over time, fewer and fewer teachers were identified as low-performing in most of the sites. The study found some evidence that this shift may have been due to increasingly generous ratings on subjective parts of the evaluation like classroom observations, rather than an actual improvement in teaching.

Past, independent research has shown that **principals rate nearly all teachers as "effective,"** but when principals are asked their opinions of teachers in confidence, they're much more likely to give harsh ratings. Principals point to the need for positive relationships with their staff members, concerns about teacher turnover, and a lack of time as potential reasons for the score inflation.

The RAND study echoes some of those findings: School leaders told researchers they would rather help teachers improve instead of dismissing them. The study suggests that because the initiative had sites use

evaluation results as the basis for tenure and dismissal decisions, principals might have avoided giving low observation ratings.

However, Dwyer said Hillsborough schools, at least, had safeguards in place to prevent that scenario from happening.

The rigorous nature of the observation rubric recommended by Gates also added a considerable time burden on administrators. Stecher said that if the evaluation scores were to be used in personnel decisions, the observations had to be rigorous and reliable—for example, a principal might need to observe a teacher for a whole hour, four times a year.

But shorter classroom drop-ins might provide helpful, more immediate feedback for a teacher, which school leaders were more interested in. Over time, some of the sites reduced the length and frequency of the observations to free up more time for administrators and to better support teacher improvement, which was not the original intent of the initiative.

“There was a real tension between using these measures for accountability purposes ... and using them for improvement tools,” Stecher said. “I don’t think any of the sites negotiated that tension perfectly, and I think it’s a difficult one for others to do as well.”

Another challenge for districts was that they didn’t have successful models on which to base some of their reforms, particularly evaluation-linked professional-development systems, Stecher said. That made it harder for sites to develop new, innovative practices.

“The big takeaway for me from this work is that maybe it might be even harder to go into existing systems with all of their routines and job descriptions and contracts and cultures and just change them in terms of their approach to evaluation and professional practice than we understood,” said Frederick M. Hess, the director of education policy studies at the American Enterprise Institute. (Hess also [authors an opinion blog at edweek.org](#).)

“It’s not just about applying big gobs of money and consulting and encouragement and even policy changes, it’s about execution,” Hess added. “Execution is not about goals and vision; execution is about dozens of very small decisions made everyday.”

The Legacy of Reform

Although student performance largely did not improve enough to meet the initiative’s goals, the study did find some positive consequences of the reforms. For instance, most teachers surveyed in all the districts said they have become more reflective about their teaching and have made changes to their instruction as a result of the evaluation system.

A spokeswoman for Hillsborough schools said in an email that it will take more time to gauge how the Gates-led practices affected student achievement, and that changes to state testing may have skewed the results. The district’s graduation rate has reached an all-time high, she said, attributing the increase to stronger instruction.

School sites will keep some of the practices they used during the initiative, even without ongoing Gates support. For instance, all sites will continue to use multiple measures for teacher evaluations, and most will continue to incorporate observation scores and student achievement growth into one composite measure that will identify low-performing teachers.

Of course, not all of that is by choice: “A lot of the stuff that was implemented are things that are required by law now,” Dwyer said. “And at the core, it is a good evaluation system.”

The Gates partnerships, while not entirely successful, will inform future research and initiatives, researchers and analysts said.

“One of the things philanthropy can and should do is experiment and let us learn about what works,” Hess said. “It was an expensive experiment, but it was a reasonable hypothesis. ... For good or bad, we’ve learned a lot. Not only about teacher evaluation, but about this approach to trying to change how school systems work.”

WEB ONLY