

Intensive Partnerships *for* Effective Teaching Enhanced How Teachers Are Evaluated But Had Little Effect *on* Student Outcomes

The initiative involved three school districts and four charter management organizations (CMOs).



The seven sites that participated in the Intensive Partnership initiative agreed to develop a robust measure of teaching effectiveness, including a structured way to observe and assess classroom teaching. They were then to use the information on effectiveness in conjunction with new or revised policies designed to do the following:

- Improve staffing by
 - recruiting and hiring teachers with high potential for effectiveness
 - placing the most-effective teachers in schools with the most LIM students
 - retaining effective teachers and removing ineffective ones.
- Identify teachers' strengths and weaknesses and provide effectiveness-linked training and support.
- Use compensation and new career roles to encourage effective teachers to stay in teaching and provide support for other teachers.

Over the course of the initiative from 2009 through 2016, spending on the reforms across the seven sites totaled \$597 million: \$237 million in grants from the Gates Foundation and the remainder primarily from each site's general fund, federal grants, and other local sources.

The Intensive Partnerships for Effective Teaching initiative, designed and funded by the Bill & Melinda Gates Foundation, was a multiyear effort aimed at increasing students' access to effective teaching and, as a result, improving student outcomes. It focused particularly on high school graduation and college attendance among low-income minority (LIM) students.

The foundation asked a team of researchers from the RAND Corporation and the American Institutes for Research to evaluate whether the initiative improved teaching effectiveness and student outcomes.

The team found that, despite the sites' efforts and considerable resources, the initiative failed to achieve its goals for improved student achievement and graduation, although the sites did implement improved measures of teaching effectiveness. With minor exceptions, student achievement, LIM students' access to effective teaching, and graduation rates in the participating districts and CMOs were not dramatically better than at similar sites that did not participate in the initiative.

This brief, based on a longer final report, summarizes the findings of the team's evaluation and offers some possible reasons the Intensive Partnership initiative did not achieve its goals for students.

Measures of Teaching Effectiveness

As determined by the site-developed measures, almost all teachers were considered to be effective.

Each participating site designed a teacher evaluation system that included at least two factors in a composite score: (1) rubric-based ratings based on classroom observations and (2) a measure of student achievement growth.

According to the measures the sites developed and the thresholds they set for teaching effectiveness, almost all teachers were deemed effective (see Figure 1). Over time, more and more teachers were rated in the top effectiveness categories, and fewer and fewer were rated as ineffective. By the end of the initiative, just 1 to 2 percent were classified as ineffective in most of the sites. This might reflect actual improvement in teaching effectiveness, but there is some evidence that it is due to other factors, such as increasingly generous ratings on subjective components (e.g., classroom observations).

The evaluation system raised some practical challenges that sites addressed in different ways. One was that the observations placed a burden on principals' time, so some sites reduced the length or frequency of classroom observations or allowed other administrators to conduct them. Another was that many teachers did not receive individual scores for their contribution to student achievement because there were no standardized tests in their subjects or grade levels. Some sites handled this by assigning a school-level average score to those teachers; others adopted alternative ways to measure student growth.

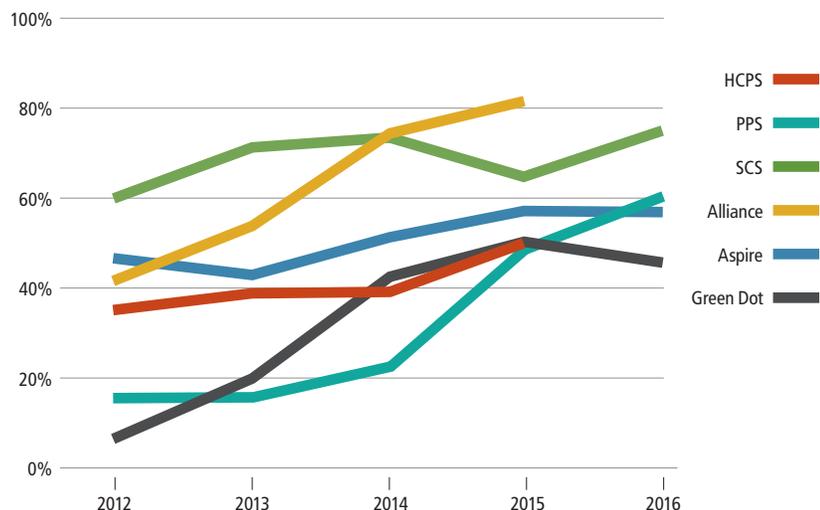
Despite some concerns about fairness, surveys that the team administered found that the majority of teachers thought that the evaluation measures were a valid measure of their effectiveness as teachers, particularly the classroom-observation component. Furthermore, most teachers thought that the evaluation system had helped them improve their teaching.

Almost all teachers were considered to be effective

NOTES: HCPS and Alliance did not provide effectiveness ratings for 2016. PUC Schools has not used overall effectiveness ratings since 2013 and thus is omitted.

Figure shows the percentage of teachers in the top two (of five) levels of effectiveness in each site except PPS, for which it shows the percentage in the top level (of four).

FIGURE 1
Percentage of Teachers Rated in the Top Effectiveness Categories, over Time



Effectiveness-Based Staffing Policies

The initiative had little effect on the retention of effective teachers, but it did increase the rate of departure of ineffective teachers.

The sites made efforts to retain effective teachers, including offering additional compensation and career opportunities based on effectiveness. However, in the end, effective teachers were no more likely to be retained after the initiative than before it.

On the other hand, ineffective teachers were more likely than before to depart from the sites. Across the sites for which data were available, about 1 percent of teachers were dismissed for poor performance in the 2015–2016 school year. Sites dismissed few teachers at least partly because their evaluation systems identified very few poor performers; however, the likelihood that those identified as poor performers would leave the site—whether voluntarily or involuntarily—increased during the initiative.

The three districts set specific criteria based on their new evaluation systems to identify low-performing teachers who might be denied tenure, placed on improvement plans, or considered for dismissal or nonrenewal of their contracts. The CMOs (which do not offer tenure) did not establish specific criteria to identify low performers but did take teacher evaluation results into account when considering improvement plans or contract renewal.

The sites also had to deal with the potentially conflicting goals of using measures of teaching effectiveness for dismissing low-performing teachers and using them to help teachers improve. In general, they tended to favor trying to help teachers improve rather than dismissing them.

All the sites modified their recruitment and hiring policies somewhat during the initiative—for example, by facilitating hiring in hard-to-staff schools or developing partnerships with local colleges. However, the researchers found little evidence that the new policies led to the hiring of more-effective teachers. Although school leaders generally thought that hiring processes worked fairly well, the sites still had difficulty attracting effective teachers to high-need schools, and persistent teacher turnover was a particular problem for the CMOs.



At our school, I don't know of anybody [who] has been transitioned out solely based on their evaluation. It has become evident to them as a person that they're not where they should be, they shouldn't be doing this teaching, and they've chosen to leave. I don't think we've had to force . . . anyone [to] leave based on numbers.

—TEACHER

Effectiveness-Based Professional-Development and Support Policies

Evaluation-linked professional development (PD) and support were difficult to achieve.

All the sites offered multiple types of PD, including coaching, workshops, school-based teacher collaboration, and online and video resources. However, the sites struggled to figure out how to organize this training and support to address individual teachers' identified needs.

We have not been able to move in our district from [implementing] the evaluation component to, now, “How do we deliver the kind of professional development that that teacher needs, based on what we now know about their practice, so they can improve on their practice and become better instructors of students?”

—CENTRAL-OFFICE ADMINISTRATOR

One possibility is that scores and feedback from the measures of teaching effectiveness might not have been detailed enough to support specific suggestions for customized PD, and existing PD systems might not have been flexible enough to provide such customization. Also, there were few existing models of evaluation-linked PD that the sites could easily adopt, and sites lacked the capacity to develop and implement new models themselves.

Most school leaders said that they suggested PD and support based on teachers' evaluation results, but the sites generally did not require teachers to participate, monitor their participation, or examine whether participants' teaching effectiveness improved as a result. In addition, some also found it difficult to develop a coherent system of PD offerings.

Teachers in all the sites generally believed that the PD activities in which they participated were useful for improving student learning. Most teachers had access to some form of coaching, on which the sites often relied to individualize PD, and the percentage of teachers with access to coaching increased over time. Teachers with lower ratings were more likely than higher-rated teachers to report receiving individualized coaching or mentoring, but they were generally no more likely than higher-rated teachers to say that the support they received had helped them.



Effectiveness-Based Compensation and Career-Ladder Policies

Some compensation and career-ladder policies were enacted to retain effective teachers, but they were not as extensive as envisioned, did not always follow best practices, and were not necessarily incentives about which teachers cared.

All seven participating sites implemented effectiveness-based compensation reforms, which varied in terms of timing, eligibility criteria, dollar amounts, and the proportion of teachers earning additional compensation. Teachers generally endorsed the idea of additional compensation for outstanding teaching, but (except in two of the CMOs) most reported that their sites' compensation systems did not motivate them to improve their teaching. See Figure 2.

All seven sites also introduced specialized roles, with additional pay, open to effective teachers who accepted additional responsibility to provide instructional or curricular support to other teachers. However, none of the sites implemented career ladders, in which specialized roles come with sequential steps and growing responsibility, like the initiative sponsors had envisioned. The districts and CMOs took somewhat different approaches to creating specialized roles for teachers. The districts created a few positions that focused on coaching and mentoring new teachers in struggling schools, while the CMOs created more positions with a wider range of duties as needs shifted over time.

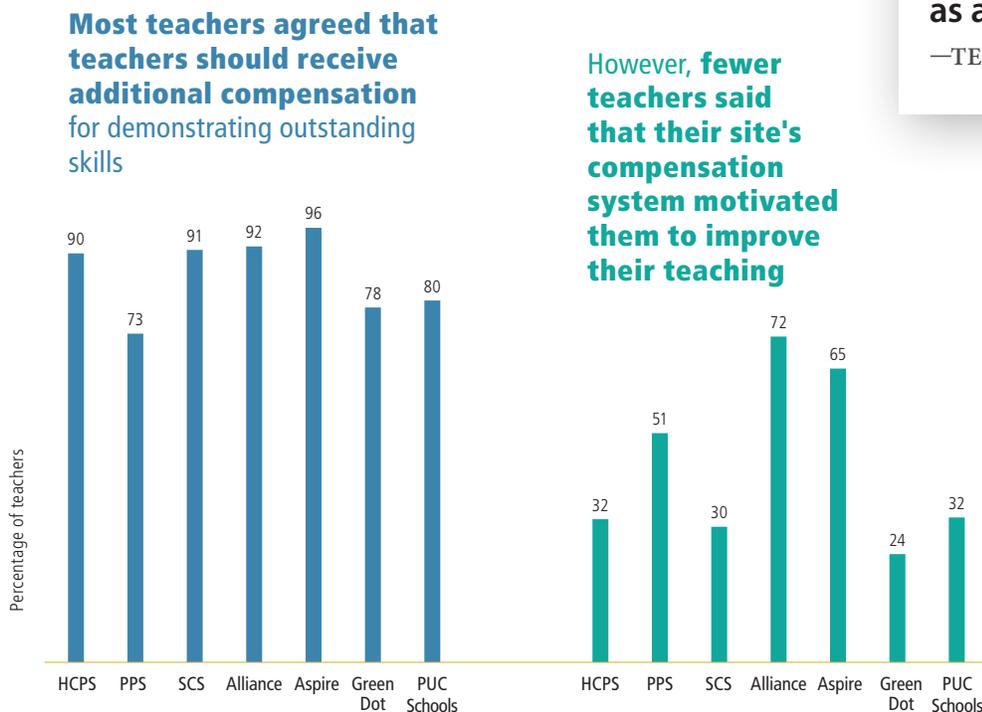


[The bonus] is really great; who's going to complain? But we're not in the job because it's a lucrative field. We do it because we love what we do. It's nice to have that fiscal validation, but it's not the deal-breaker for motivating whether I'm going to stay as a teacher or not.

—TEACHER

FIGURE 2

Percentage of Teachers Agreeing Strongly or Somewhat with Statements About Their Sites' Compensation Systems



Teaching Effectiveness, Access to Effective Teachers, and Student Outcomes

The initiative did not achieve its goals of increasing teaching effectiveness overall, improving access to effective teaching for LIM students, or boosting student outcomes.

The analysis found little evidence that teaching effectiveness improved as a result of the initiative. This was true whether teaching effectiveness was measured by the sites' own composite measures or by an independently calculated measure. The researchers also looked for an increase in the teaching effectiveness of newly hired teachers but did not find evidence of one. As mentioned, the departure rate for the least effective teachers increased at some sites, although this success was not sufficient to noticeably improve the average teaching effectiveness of those sites.

At the beginning of the initiative, LIM students had roughly the same access to effective teaching as all students, and their access had not improved by the end of the initiative. In addition, their achievement and graduation rates appeared no different from those of their peers in similar schools that did not participate in the initiative (see table). Similarly, the analyses of test results and graduation rates for students overall showed no evidence of the initiative having a widespread positive impact in most sites and grade ranges. However, the initiative did have a significant positive effect in high school English in the CMOs and PPS but a significant negative effect in grade 3–8 mathematics in the CMOs.

Two caveats should be considered in interpreting these results. First, teacher-evaluation mandates with consequences were enacted in three of the four states at the same time as the initiative, so the comparison sites and the sites participating in the initiative were exposed to some of the same types of new policies. The team's impact estimates reveal the extent to which the initiative improved student outcomes over and above these statewide efforts. Second, it is possible that the reforms simply require more time to take effect, so the research team is monitoring student outcomes for two additional years.



The initiative's goal of **boosting student outcomes** was not achieved

+ Positive effect

– Negative effect

□ No effect

THE INITIATIVE'S ESTIMATED IMPACT ON STUDENT ACHIEVEMENT

2014–2015, by grade span and site

SITE	GRADES 3–8		HIGH SCHOOL	
	Mathematics	Reading	Mathematics	Reading
HCPS			N/A	
PPS			N/A	+
SCS			N/A	N/A
CMOs	–		N/A	+

NOTE: The researchers could not estimate the impact on high school mathematics because students did not take the same secondary mathematics tests. N/A = not applicable.

Statistical significance measured at $p < 0.05$.

Why Didn't the Initiative Achieve Its Goals?

The initiative had greater success implementing measures of teaching effectiveness than improving student outcomes.

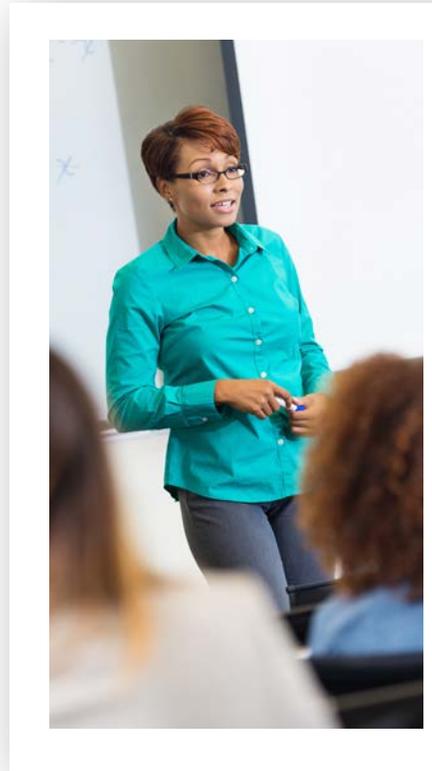
A favorite saying in the educational measurement community is that one does not fatten a hog by weighing it. In the end, the sites were better at implementing measures of teaching effectiveness than at using them to improve student outcomes. The RAND/American Institutes for Research evaluation of the Intensive Partnerships for Effective Teaching initiative does not explain why the desired student outcomes were not achieved, but, informed by observations of the sites over the past seven years, the team can speculate about potential explanations:

Implementation was incomplete and lacked successful models. It is possible that the new policies were not implemented with sufficient quality, intensity, or duration to achieve their potential full effect. None of the main policy levers—staffing, PD, or compensation and career ladders—was implemented fully or as initially envisioned. Incomplete implementation might have been due, in part, to a lack of successful models, which gave the sites little to go on, and the sites might have lacked the capacity to develop such models on their own.

Using teacher-evaluation measures toward different goals might create a conflict. The sites found it difficult to navigate the underlying tension between using teacher evaluation to help teachers improve and using it to make high-stakes decisions about compensation, tenure, and dismissal.

State and local contexts changed. Some local and state changes in context could have interfered with the sites' abilities to fully implement the reforms. During the initiative, all four states changed their statewide tests, two squeezed education budgets, one school district merged with another, and some districts had turnover in top leadership. In addition, new teacher evaluation measures with consequences were enacted in three of the states, and these policy changes affected both the IP sites and the schools we used as a comparison. Thus, our impact estimates reveal how well the IP initiative improved student outcomes *over and above* these statewide efforts.

Despite the initiative's failure to improve student outcomes, the sites still use many of the policies, either because they found them valuable or because state law or regulation now requires them. In particular, the sites continue to incorporate systematic teacher evaluation into regular practice and have kept many new recruitment and hiring policies.





To ensure that all students graduate from high school prepared to succeed in and complete college, school systems must produce [large] gains in student achievement and attainment Fulfilling this goal will require shifts in beliefs, policies, and actions that ultimately empower a school system's single largest asset—its corps of teachers—to deliver the highest-quality instruction in every school, in every classroom, and to every student.

—Bill & Melinda Gates Foundation, *Intensive Partnerships to Empower Effective Teachers: Invitation-Only Request for Proposals*, Seattle, Wash., 2009

KEY FINDINGS

The sites succeeded in implementing measures of effectiveness to evaluate teachers, but almost all teachers were rated effective or above.

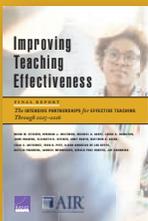
By the end of the initiative, ineffective teachers were more likely than before to leave teaching, but effective teachers were no more likely to remain teaching.

Although the sites made use of the measures to some degree in human-resource decisions (for example, compensation and dismissal), the sites did not draw on the measures to the extent anticipated—for example, to inform and provide effectiveness-linked professional development to teachers.

With minor exceptions, student achievement, access to effective teaching, and dropout rates in the participating sites were not dramatically better than they were for similar sites that did not participate in the initiative.

The initiative did not generally increase low-income minority (LIM) students' access to more-effective teaching.

Overall, the initiative did not achieve its goal of increased student achievement and graduation.



This brief describes work funded by the Bill & Melinda Gates Foundation done in RAND Education and Labor documented in *Improving Teaching Effectiveness: Final Report: The Intensive Partnerships for Effective Teaching Through 2015–2016*, by Brian M. Stecher, Deborah J. Holtzman, Michael S. Garett, Laura S. Hamilton, John Engberg, Elizabeth D. Steiner, Abby Robyn, Matthew D. Baird, Italo A. Gutierrez, Evan D. Peet, Iliana Brodziak de los Reyes, Kaitlin Fronberg, Gabriel Weinberger, Gerald Paul Hunter, and Jay Chambers, RR-2242-BMGF, 2018 (available at www.rand.org/t/RR2242).

The brief was updated in 2019 to include additional information discussed in RR-2242/2-BMGF, *Improving Teaching Effectiveness: The Intensive Partnerships for Effective Teaching Through 2015–2016: Addendum*, by Michael S. Garett, Deborah J. Holtzman, Brian M. Stecher, and John Engberg. To view this brief online, visit www.rand.org/t/RB10009-1.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

This work is licensed under a Creative Commons Attribution 4.0 International License. All users of the publication are permitted to copy and redistribute the material in any medium or format and transform and build upon the material, including for any purpose (including commercial) without further permission or fees being required.

Learn more at gatesfoundation.org.

www.rand.org