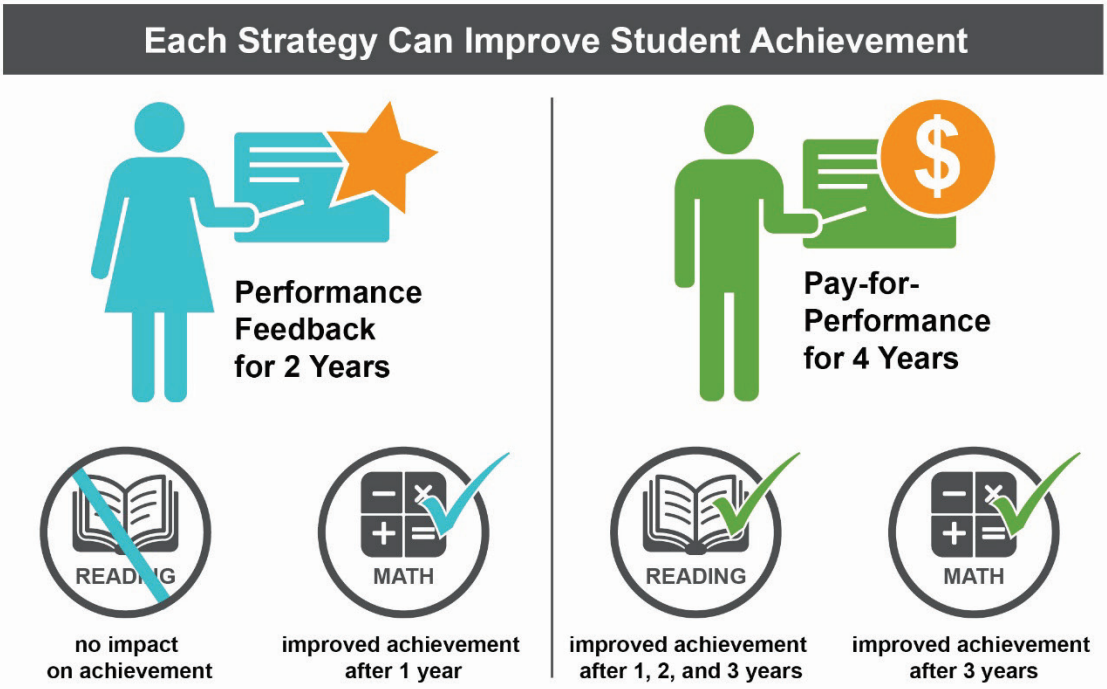
ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

PROMOTING EDUCATOR EFFECTIVENESS

THE EFFECTS OF TWO KEY STRATEGIES



Recent studies from the Institute of Education Sciences evaluated different strategies supported by the **Teacher Incentive Fund**. The strategies focused on increasing the effectiveness of teachers and principals as a way to improve student achievement.

Providing educators with feedback on their performance (performance feedback) for two years improved students' math achievement after the first year by about 4 weeks of learning. The cumulative effect after 2 years of implementation was similar in magnitude but not statistically significant. The effect on reading in both years was positive but not statistically significant.

Providing educators with bonuses based on their performance (pay-for-performance) for four years improved students' reading achievement after 1, 2, and 3 years of implementation and students' math achievement after 3 years. After each of those periods of implementation, the cumulative effect amounted to about 3 to 4 weeks of learning.

Having a more effective teacher or principal can substantially improve students' academic outcomes.[1] The Teacher Incentive Fund (TIF) program, established in 2006, provided competitive grants to help states and districts implement a multi-strategy approach to enhancing educator effectiveness. TIF grantees were required to measure educator performance and use this information to make decisions about the support and compensation they provide for educators. The 2015 reauthorization of the Elementary and Secondary Education Act replaced TIF with the Teacher and School Leaders (TSL) Incentive Grants program. This program, like the TIF program, provides grants to eligible entities to develop, implement, improve, or expand performance-based compensation systems and human capital management systems in schools.

This brief brings together the findings of two studies from the Institute of Education Sciences (IES) that examined specific strategies supported by TIF. One study,[2] conducted over two years, examined the effects of using research-based performance measures to evaluate educators and provide them with feedback—a strategy referred to here as *performance feedback*. The second study,[3] conducted over four years, evaluated the effects of offering bonuses to educators based on their performance ratings—a strategy referred to as *pay-for-performance*. Although no single large-scale study has evaluated the effects of a full, comprehensive program, like those supported by TIF and TSL, these studies can provide insight on the potential effects of two strategies that are prominent in such programs. In addition, evidence from both studies suggests areas of potential improvement for programs that support or expand human capital initiatives in schools and highlights potential avenues for future research.

Key Findings

Providing educators with performance feedback and offering pay-for-performance bonuses can improve student achievement. Across years of implementation, each tested strategy sometimes had a positive cumulative impact on students' math or reading achievement.

Features of the measures or programs and on-the-ground implementation may limit the effectiveness of the program strategies. Both studies provided evidence that the policy as implemented differed in some ways from the approach envisioned. For example, in the pay-for-performance study, about 40 percent of the teachers were unaware they were eligible to receive a performance bonus.

Educators can be receptive to some of the evaluation and compensation strategies supported by TIF and TSL. Sixty-five to 84 percent of the educators reported being satisfied with the feedback they received on their practices. In addition, pay-for-performance ultimately led to improvements in teachers' satisfaction with some aspects of their jobs.

The Teacher Incentive Fund program

TIF grants provided federal support for programs that comprehensively reform how states and districts measure educator performance, provide feedback, and compensate educators. Across four rounds of TIF grants (in 2006, 2007, 2010, and 2012), the U.S. Department of Education awarded about $1.8 billion to help states and districts undertake such reforms.[4]

The 2010 TIF grantees were required to implement the strategies covered by the performance feedback and pay-for-performance studies. In particular, the 2010 TIF grants required that grantees' TIF programs include the following:

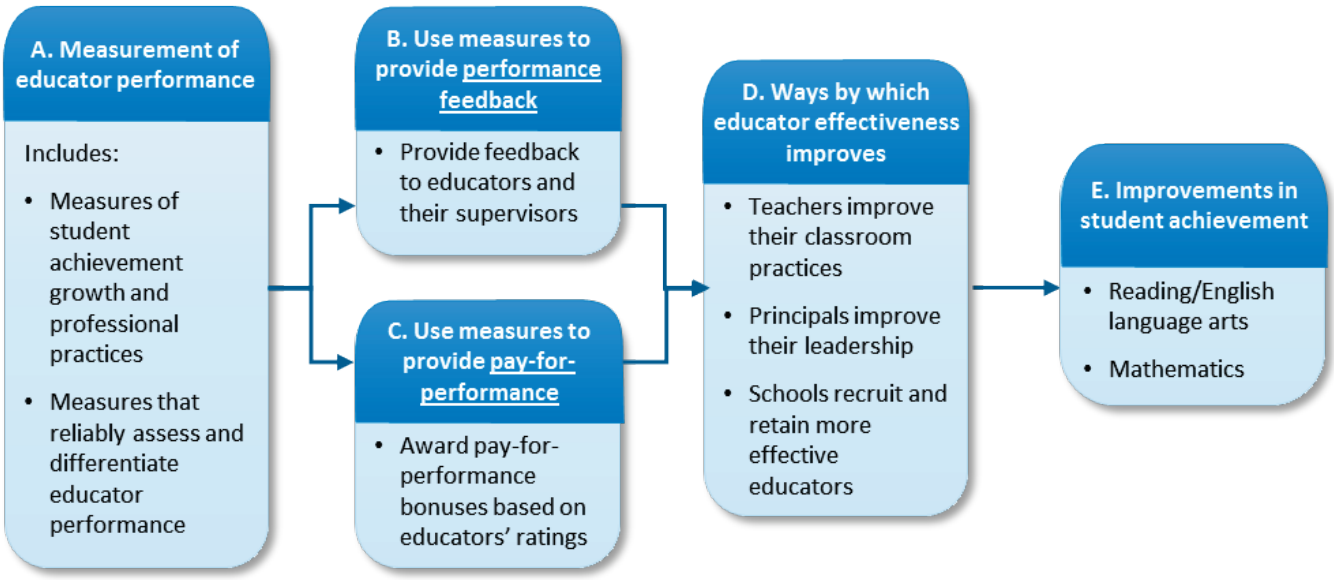
- **Measures of educator effectiveness** based on student achievement growth and multiple observations of classroom or leadership practices per year by trained observers
- Professional development to help educators understand how they were evaluated and to provide **feedback based on educators' performance ratings** to improve their practices
- **Additional pay opportunities** for effective educators who take on additional roles or responsibilities, such as a master or mentor teacher
- **Pay-for-performance bonuses** that are based on how well educators perform on the measures of educator effectiveness and that should be substantial in size, differentiated, and challenging to earn

The first two strategies were the focus of the performance feedback study; the pay-for-performance study focused on the last strategy.

These strategies share a theory of action for how measuring educators' performance to provide educators with feedback and offering performance bonuses might lead to improved student achievement (Exhibit 1). If a district adopts performance measures, based in part on student growth, that can distinguish between lower and higher performing educators (Box A), the district can use this performance information for different purposes. For example, the district can use this information to provide feedback to educators and their supervisors on their performance. This appears in Box B and is the focus of the performance feedback study. Alternatively, the district can use the performance information to award performance bonuses. This appears in Box C and is the focus of the pay-for-performance study.

Using the information for either purpose could lead to improvements in educator effectiveness in three ways (Box D). First, teachers may learn about aspects of their classroom practices that could be improved, or become more motivated to improve their practices, and therefore improve their effectiveness. Second, principals could learn how to improve their leadership skills or be more motivated to do so. Third, providing performance feedback or offering pay-for-performance bonuses may help retain or attract more effective educators to work in districts and schools where their effectiveness is recognized and rewarded. Any of these effects could result in improved student achievement (Box E).

Exhibit 1. Theory of action for strategies that use measures of educators' performance to provide educators with feedback and offer performance bonuses



For these improvements in educator effectiveness and student achievement to be realized, the ways in which districts design and implement performance feedback and performance bonuses, as well as educators' attitudes toward the strategies, may be important. For example, according to proponents of performance feedback and pay-for-performance, the performance measures should distinguish between an educator's strengths and weaknesses.[5] The measures should also capture dimensions of practice that are important to improving student achievement.[6] In addition, pay-for-performance bonuses should be structured to motivate educators to change their behaviors, and educators must be aware of and understand the bonus program.[7] Furthermore, it may be important to gain educators' support for the initiatives, as lack of support could decrease educators' job satisfaction and, as a result, lower their effectiveness.[8] If the performance measures or the bonus programs fall short of these goals, their ability to improve student achievement may diminish.

Two random assignment studies

To learn more about the effectiveness of these two strategies promoted by the TIF and TSL programs, IES sponsored two random assignment studies—one focused on providing performance feedback and the other on pay-for-performance bonuses. In each study, elementary and middle schools were assigned randomly—that is, completely by chance—to implement the strategy or not. Each study compared the students in these two groups in the spring of each year. Because of the use of random assignment, differences in student achievement between the two groups after a strategy was implemented can be attributed to the strategy. Each study calculated these differences based on students tested in the study schools at the end of each school year, regardless of how long the students had been enrolled. Although the studies both focused on strategies promoted by TIF and TSL programs, they targeted different types of schools, used different measures, and used different feedback measures. Exhibit 2 summarizes these and other key aspects of the two studies.

The performance feedback study tested the strategy of evaluating educators with performance measures and using the performance information to provide feedback. Teachers and principals received the additional

feedback for two years. The observation instruments used to assess teachers' classroom practices measured performance on multiple dimensions (e.g., *Behavior management* and *Regard for student perspectives*). Using such instruments, teachers received ratings as well as written and oral feedback on their classroom practices four times a year, each time including ratings on each measured dimension as well as an overall rating. Teachers additionally received scores on their contribution to student achievement growth (known as value added) once a year. Principals received ratings and oral feedback on their leadership twice a year using a 360-degree survey tool. Each time leadership was assessed, the principal received ratings on multiple dimensions of his or her leadership (e.g., *High standards for student learning* and *Culture of learning and professional behavior*) as well as an overall rating. Although the strategies in the performance feedback study were not directly funded by TIF or TSL, they were similar to the performance feedback strategies that the programs have promoted. In the study, no formal consequences were attached to the performance measures—for example, they were not used for employment or tenure decisions.[9] Instead, the measures were used to provide educators and their supervisors with information about performance. Such information might identify educators who need support and indicate areas for improvement.

The pay-for-performance study examined the strategy of offering pay-for-performance bonuses to teachers and principals within a comprehensive evaluation system supported by the 2010 TIF grants. For four years, all schools in the pay-for-performance study were required to implement strategies similar to those tested by the performance feedback study—for example, evaluating teachers based on student achievement growth and formal observations, and providing them with feedback. However, to examine the effects of pay-for-performance, the study team assigned half of the schools to offer bonuses based on those measures, and the other half not to offer performance bonuses. Districts in the study received examples of how to design bonuses so that they might successfully motivate educators to improve their practices (e.g., making bonuses challenging to earn). The study examined the extent to which actual bonuses incorporated the guidance suggested and how well educators understood the bonus program.

Exhibit 2. Summary of IES studies of performance feedback and pay-for-performance

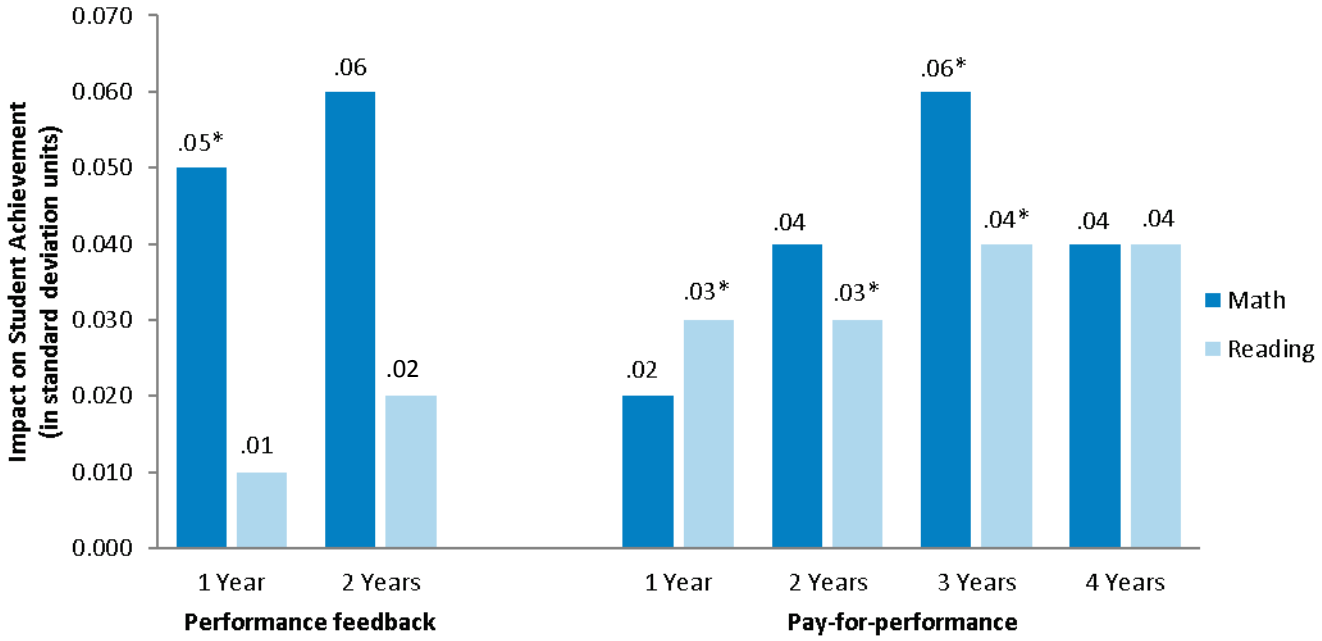
	The Performance Feedback Study	The Pay-for-Performance Study
District policy context	Districts with evaluation system requirements that were less objective and intensive than the performance feedback that was tested	Districts that applied for and received a 2010 TIF grant, including a subset of districts that received an "evaluation" grant in which they agreed to participate in a random assignment study of pay-for-performance
Strategy and how it was tested	Treatment and control schools continued to implement their existing evaluation systems. Treatment schools additionally: • Evaluated teachers' classroom practices and provided feedback 4 times per year • Evaluated teachers' contributions to student achievement growth (value added) and provided a report as feedback once per year • Evaluated principals' leadership and provided feedback 2 times per year	Treatment and control schools were expected to implement the components required for TIF grants, except that only treatment schools implemented pay-for-performance bonuses for teachers and principals. Grantees were encouraged to make bonuses: • Substantial in size • Differentiated • Challenging to earn
Duration and school years during which strategy was implemented	Two school years: 2012–13 and 2013–14	Four school years: 2011–12, 2012–13, 2013–14, and 2014–15
Types of schools and teachers	• Mix of low-, medium-, and high-poverty schools • Grades: 4–8 • Teachers of math or reading	• High-poverty schools (at least 50 percent of students eligible for free or reduced-price lunch) • Grades: 3–8 • All teachers in participating schools
Number of study participants in the last year of the study	29,385 students, 1,038 teachers, 126 schools, 8 districts, 5 states	38,939 students, 3,544 teachers, 131 schools, 10 districts, 7 states
Student outcomes	End-of-year reading and math scores of students currently attending study schools	End-of-year reading and math scores of students currently attending study schools

The two studies examined different strategies but both have the common objective of improving educator effectiveness and, ultimately, student achievement. Together, the findings from both studies can therefore provide a more complete picture of the potential effects that a comprehensive human capital management initiative might have on student achievement.

Providing educators with performance feedback and offering pay-for-performance bonuses can improve student achievement

Across the two years of the performance feedback study and the four years of the pay-for-performance study, there were sometimes positive cumulative impacts on students' math or reading achievement (Exhibit 3).

Exhibit 3. Impact of performance feedback and pay-for-performance on student achievement, by number of years of implementation



Sources: Student administrative data. Performance feedback: *N* = 29,874 students for math after Year 1; *N* = 29,995 students for math after Year 2; *N* = 28,492 for reading after Year 1; *N* = 29,385 students for reading after Year 2. Pay-for-performance: *N* = 40,535 students for math after Year 1; *N* = 40,454 students for math after Year 2; *N* = 39,770 for math after Year 3; *N* = 38,939 for math after Year 4; *N* = 40,256 students for reading after Year 1; *N* = 40,122 students for reading after Year 2; *N* = 39,538 for reading after Year 3; *N* = 38,929 for reading after Year 4 reading.

Figure reads: At the end of Year 1, the impact of performance feedback on students' math achievement was .05 standard deviation units; the impact on students' reading achievement was .01 standard deviation units.

*The impact on student achievement is statistically significant at the .05 level, two-tailed test.

The performance feedback study found a positive impact on students' math achievement after the first year of the two-year implementation period, amounting to about four weeks of learning. After two years, the impact on math achievement was similar in magnitude but not statistically significant. There was no impact on reading achievement after one or two years.

The pay-for-performance study found positive impacts on students' math and reading achievement. For math, from the end of the second year onward, schools that offered pay-for-performance bonuses had higher achievement than schools that did not, but the difference was statistically significant in only one year. For reading, from the end of the first year onward, schools that offered pay-for-performance bonuses also had

higher achievement than schools that did not, and the difference was statistically significant in three out of four years. In both subjects, these impacts amounted to about three to four weeks of learning.

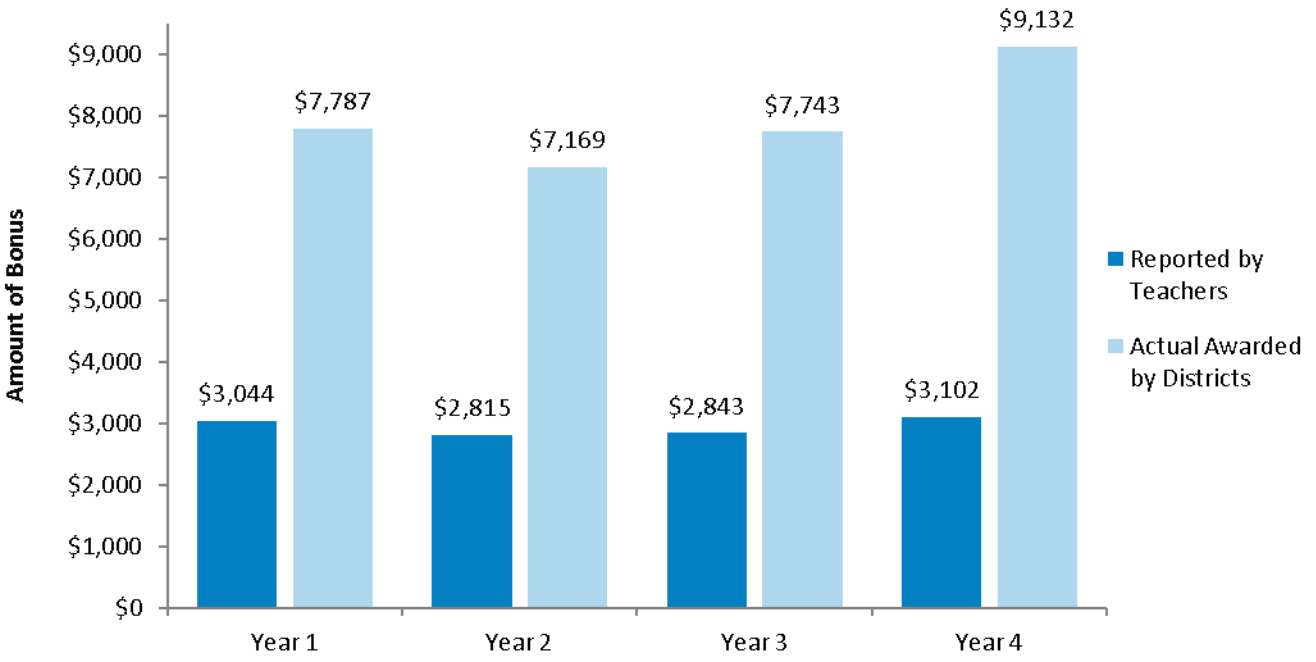
Features of the measures or programs and on-the-ground implementation may limit the effectiveness of program strategies

Fluctuations in performance ratings limited what could be learned for improving practice. In theory, measures of classroom practice and principal leadership should distinguish between lower- and higher-performing educators. Measures should also distinguish between different dimensions of an educator's performance (e.g., between a teacher's ratings on different dimensions of classroom practice such as *Behavior management* and *Regard for student perspectives*). The ratings could thus help identify educators in need of support, as well as the practices an educator should improve. Analysis of the ratings provided to educators in the performance feedback study found mixed results. For example, averaging together a teacher's overall ratings for classroom practice across the four observations in a year provided some reliable information to identify teachers who needed support. However, fluctuations in a teacher's ratings across observations limited how much one could learn from a single observation.[10] Similarly, the ratings did not provide a consistent message about an educator's relative performance on different aspects of practice. For example, while most teachers received ratings of classroom practice that differed across dimensions, the differences between dimension ratings had limited reliability to identify what a teacher most needed to improve, even when averaged across the four observations in a year.[11]

The ratings and bonuses teachers received may have limited their motivation to improve. In the performance feedback study, the ratings of teachers' contribution to student achievement growth were spread out across the scale as intended, but the ratings of classroom practices placed 85 percent of teachers or more in the top half of the rating scale, limiting the number of teachers who may have received a message that they need to improve. In the pay-for-performance study, bonuses were easy enough to earn that about 70 percent of teachers got a bonus, which may have led teachers to believe that they were already performing well.

Many educators were unaware of key program information. In the performance feedback study, teachers and principals received feedback on classroom practices and principal leadership with the intended frequency, but some did not access the first wave of scores on teachers' contribution to student achievement growth. That year, only 39 percent of the teachers with value-added scores and 40 percent of principals viewed those scores, limiting the potential impact of this performance information. Similarly, in the pay-for-performance study, from the second year of implementation forward about 40 percent of teachers were unaware they were eligible for a performance bonus, and therefore could not have been motivated by it. In addition, teachers underestimated how much they could earn from a performance bonus, which again may have reduced the impact of the bonuses (Exhibit 4). Across all four years of the study, teachers on average believed that the maximum bonus they could earn was no more than 40 percent of the maximum bonus that districts awarded.

Exhibit 4. Reported and actual maximum pay-for-performance bonuses for teachers in schools that offered pay-for-performance bonuses



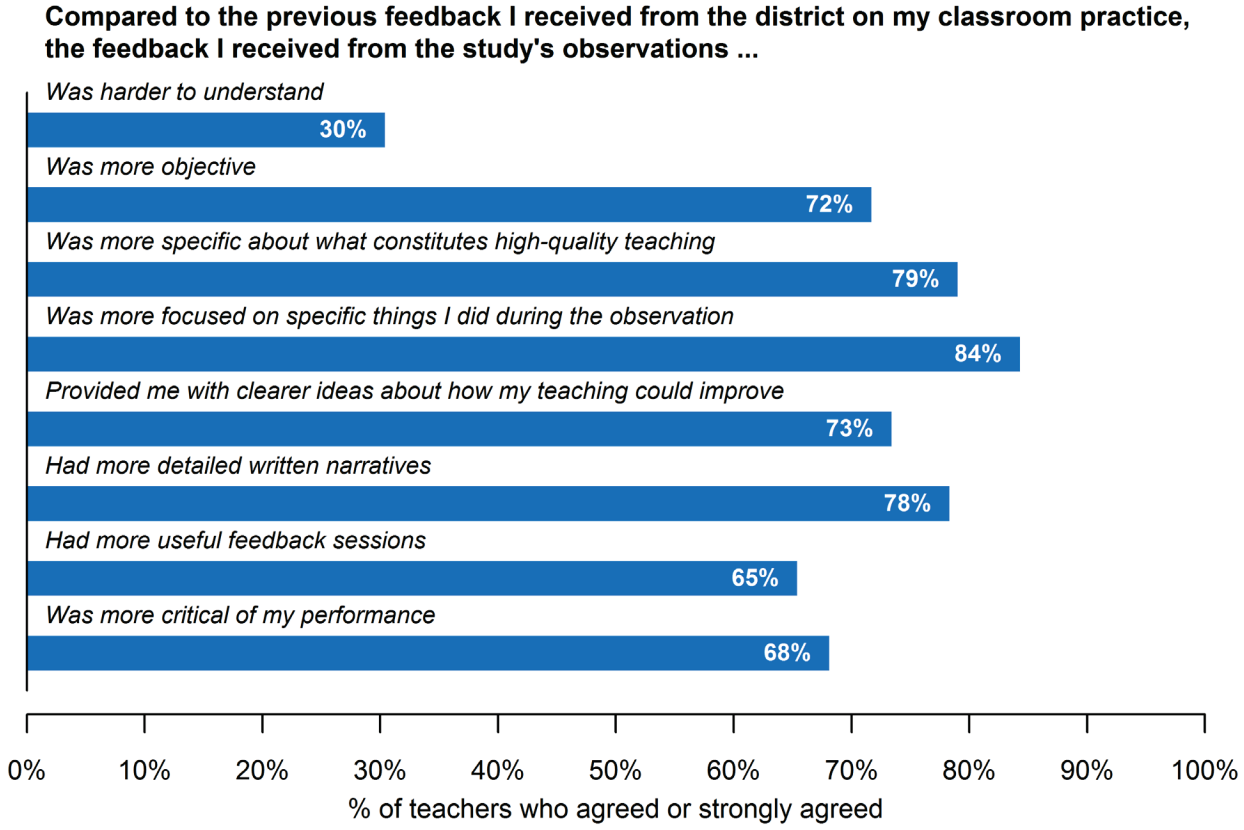
Sources: Teacher surveys, 2012, 2013, 2014, and 2015 (*N* = 218 teachers in Year 1; *N* = 229 teachers in Year 2; *N* = 229 teachers in Year 3; and *N* = 210 teachers in Year 4) and educator administrative data in the pay-for-performance study.

Figure reads: In Year 1, on average, the actual maximum pay-for-performance bonus that districts awarded to teachers was $7,787, and the maximum pay-for-performance bonus that teachers reported they could earn was $3,044.

Educators can be receptive to some of the evaluation and compensation strategies supported by TIF and TSL

Most educators reported being satisfied with the observation measures and the feedback that they received on their practices. In the performance feedback study, educators in schools that implemented the strategy were asked how the study's feedback compared with the feedback they normally received before the study. Teachers and principals for the most part said the study's feedback on educator practices was better. For example, 73 percent of teachers agreed that "the feedback provided me with clearer ideas about how to improve" (Exhibit 5). Similarly, at least two-thirds of teachers in the pay-for-performance study reported being satisfied with the use of classroom observations to assess their skills.

Exhibit 5. Percentage of teachers in the performance feedback study who agreed somewhat or strongly with each statement about the feedback based on classroom observations



Sources: Performance feedback study: Spring 2014 (Year 2) teacher survey (*N* = 320–430 teachers).

Figure reads: Thirty percent of teachers agreed somewhat or strongly with the statement "The feedback I received from the study's observations was harder to understand" compared to the feedback received prior to the intervention as part of their district's formal teacher evaluation system.

Compared to their perceptions of the feedback on classroom practices, fewer teachers held positive perceptions of the use of student achievement scores to assess their performance. In the performance feedback study, only about half of teachers (41 percent to 55 percent) held positive views on the use of value-added scores to evaluate their performance. For example, 48 percent of teachers reported that "the value-added score is a good measure of how well students learned what I taught last year." Likewise, about 42 percent of teachers reported that the value-added scores are fair to all teachers, regardless of the personal characteristics of the students they taught. The pay-for-performance study also found that after four years of TIF implementation, more teachers were satisfied with the use of classroom observations to assess their skills (about 75 percent) than with the use of student achievement scores (about 60 percent).

By the third year of implementation, pay-for-performance led to improvements in teachers' satisfaction with some aspects of their jobs. In the first two years of TIF implementation, teachers in schools that offered performance bonuses tended to report being less satisfied with some aspects of their jobs than teachers in other schools. For example, compared to teachers in other schools, teachers in schools that offered performance bonuses were less likely to be satisfied with the quality of their interactions with their colleagues (75 percent versus 81 percent after the first year of implementation) and with the use of student achievement scores to assess their performance (61 percent versus 69 percent after the second year of implementation). However, after the third year of implementation, the teachers in these schools were no longer less satisfied on any of the measured aspects of their jobs and, in fact, reported being more satisfied on some aspects. For example, after three years of implementation, teachers in schools that offered performance bonuses were more likely than teachers in other schools to be satisfied with the quality of their interactions with their colleagues (84 percent versus 79 percent) and school morale (63 percent versus 53 percent).

Concluding thoughts

This brief describes the findings of two studies that examined the impacts of implementing two key strategies for boosting educator effectiveness: (1) measuring educators' effectiveness and providing them with feedback and (2) offering performance bonuses. Individually, each strategy led to small improvements in student achievement. However, shortcomings in the design and implementation of these strategies may have prevented them from being as effective as envisioned. In particular, the performance measures used in the feedback study provided limited information to identify the aspects of practice an educator most needed to improve. The performance measures also allowed most teachers to earn high ratings and bonuses, which may have made it seem that there was little room for improvement. In addition, some educators were unaware of key information about these programs. Despite the limitations in these strategies, teachers were generally receptive to them, reporting favorable attitudes toward the feedback they received on their practices as well as increased satisfaction with some aspects of their jobs as a result of the opportunity to earn performance bonuses.

Findings from the two studies raise the question of whether adopting both performance feedback and pay-for-performance simultaneously could achieve a combined effect larger than the effect that either of these studies found individually. The performance feedback study found small, positive effects of this strategy when implemented alone, and the pay-for-performance study found small, positive effects of performance bonuses when added on to a broader system that included performance feedback. Schools that implement both strategies may therefore experience, to some degree, both of these positive effects. However, predicting the

exact magnitude of the combined effect is not possible from these two studies alone, because these two studies were not fully designed to be pieced together. In particular, the two studies occurred in different types of schools (with schools in the pay-for-performance study being more disadvantaged) and used different measures and feedback methods. Future research could evaluate the effects of a more comprehensive program that combines both strategies.

Findings from the studies of the two strategies suggest at least three possible revisions that may improve human capital initiatives and could be evaluated in future research. First, increasing the number of classroom observations per year for each teacher may make the feedback on classroom practices more useful as a guide for improving teaching. Although four observations seem sufficient to reliably identify teachers who need support, having more observations in a year, perhaps strategically targeting some teachers, would more reliably identify persistent weaknesses and thus better guide improvement for individual teachers.

Second, to indicate to educators that they have room to improve, the program could make it more difficult to earn the highest ratings or bonuses. For example, given that the vast majority of teachers receive classroom practice ratings at the top end of the scale, it is likely that ratings are inflated, either intentionally or unintentionally. In both cases, additional training could be useful. To address unintentional inflation, training on the observation rubric could emphasize the importance of giving ratings that more accurately reflect the scoring criteria. To limit intentional rating inflation, training could provide more support on how to conduct feedback sessions for those who receive a low rating. Alternatively, closely supervising how ratings are determined might lead to fewer teachers receiving inflated ratings. Another way to ensure that most educators see room to improve is to adjust criteria for earning a bonus to reward more substantial improvement.

Improving educators' awareness of program components, such as their value-added scores or the maximum bonus amounts they could receive, is a third strategy that might improve programs. Districts might explore new ways of communicating program information or simplifying the programs themselves. For example, districts could disseminate information through text-message or email reminders and provide more details on the bonus amounts the district distributed in prior years. Some ways of structuring bonus programs may also make it easier for educators to understand how much they could potentially earn. For example, educators may have a more accurate understanding of bonus amounts when they can receive specific dollar amounts for meeting each performance criterion, rather than their bonus amount being dependent on how many other educators also performed well.

Although the performance feedback and pay-for-performance studies provide insight into two important human capital management strategies, programs like TIF and TSL include other strategies that might impact student achievement. For example, TIF and TSL programs also include providing additional pay for highly effective educators to take on additional roles and responsibilities. Further research on these additional strategies would be needed to evaluate the potential impact of a more comprehensive human capital management initiative.

[1] Examples of studies showing a strong influence of teachers on student achievement growth are Aaronson, Daniel, Lisa Barrow, and William Sander. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, vol. 25, no. 1, 2007, pp. 95–135. Chetty, Raj, John Friedman, and Jonah Rockoff. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, vol. 104, no. 9, 2014,

pp. 2593–2632. Kane, Thomas, Jonah Rockoff, and Douglas Staiger. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review*, vol. 27, 2008, pp. 615–631. Rivkin, Steven, Eric Hanushek, and John Kain. "Teachers, Schools, and Academic Achievement." *Econometrica*, vol. 73, no. 2, March 2005, pp. 417–458. Rockoff, Jonah. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review: AEA Papers and Proceedings*, vol. 94, no. 2, 2004, pp. 247–252. Examples of studies showing a strong influence of principals on student achievement growth are Branch, Gregory, Eric Hanushek, and Steven Rivkin. *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals.* Working paper no. 17803. Cambridge, MA: National Bureau of Economic Research, 2012. Chiang, Hanley, Stephen Lipscomb, and Brian Gill. "Is School Value Added Indicative of Principal Quality?" *Education Finance and Policy,* vol. 11, no. 3, 2016, pp. 283–309. Chiang, Hanley, Moira McCullough, Stephen Lipscomb, and Brian Gill. *Can Student Test Scores Provide Useful Measures of School Principals' Performance?* (NCEE 2016-002). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2016. Coelli, Michael, and David Green. "Leadership Effects: School Principals and Student Outcomes." *Economics of Education Review*, vol. 31, no. 1, 2012, pp. 92–109. Dhuey, Elizabeth, and Justin Smith. "How Important are School Principals in the Production of Student Achievement?" *Canadian Journal of Economics*, vol. 47, no. 2, 2014, pp. 634–663. Grissom, Jason, Demetra Kalogrides, and Susanna Loeb. "Using Student Test Scores to Measure Principal Performance." *Educational Evaluation and Policy Analysis*, vol. 37, no. 1, pp. 3–28.

[2] Garet, Michael, Andrew Wayne, Seth Brown, Jordan Rickles, Mengli Song, and David Manzeske. *The Impact of Providing Performance Feedback to Teachers and Principals* (NCEE 2018-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2017.

[3] Chiang, Hanley, Cecilia Speroni, Mariesa Herrmann, Kristin Hallgren, Paul Burkander, and Alison Wellington. *Evaluation of the Teacher Incentive Fund: Final Report on Implementation and Impacts of Pay-for-Performance Across Four Years* (NCEE 2018-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2017.

[4] The 2015 reauthorization of the Elementary and Secondary Education Act renamed TIF the Teacher and School Leader Incentive Grants program.

[5] See, for example, Donaldson, Morgaen L., and John P. Papay. "Teacher Evaluation Reform: Policy Lessons for School Principals." *Principal's Research Review* 9, no. 5 (2014): 1–8. See also: Papay, John. "Refocusing the Debate: Assessing the Purposes and Tools of Teacher Evaluation." *Harvard Educational Review* 82, no. 1 (2012): 123–141.

[6] See generally Kane, Thomas J., and Douglas O. Staiger. "Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains" (Research Paper: MET Project). Seattle, WA: Bill & Melinda Gates Foundation, 2012.

[7] See, for example, Podgursky, Michael J., and Matthew G. Springer. "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management* 26, no. 4 (2007): 909–949.

[8] White, Bradford R., Jennifer Cowhy, W. David Stevens, and Susan E. Sporte. *Designing and Implementing the Next Generation of Teacher Evaluation Systems: Lessons Learned from Case Studies in Five Illinois Districts* (Research Brief). Chicago, IL: Consortium on Chicago School Research, 2012.

[9] There were exceptions in three of the eight districts in the study. In these districts, the observation conducted by the principal, which was one of four per teacher required as part of the treatment, counted in their official rating system if the teacher was due to be observed that year under the district's existing evaluation system.

[10] We estimated the degree to which a teacher's four-window average score was a reliable measure of the teacher's persistent classroom practice over each year, based on the variation in the four-window average scores across teachers (between-teacher variance) and the variation in a teacher's scores across the windows (within-teacher variance). These reliability estimates tell us how consistent a teacher's overall scores were over the four observation windows. Depending on assumptions about the sources of measurement error, second-year estimated reliabilities for the four-window averages were .53 to .61 and .70 to .77 for the two rubrics used. Scores from a single observation during the second year of the performance feedback study had estimated reliabilities of .33 and .51 for the two observation rubrics used. The Standards for Educational and Psychological Testing do not suggest a minimum degree of reliability, but state that the reliability evidence for a measure should be appropriate for the measure's intended use, and that a higher degree of reliability is required for uses that have more significant consequences. See American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association, 2014. For consequential personnel decisions, measures with reliabilities above .70 are often considered acceptable. See U.S. Department of Labor, Employment and Training Administration. *Testing and Assessment: A Guide to Good Practices for Workforce Investment Professionals.* Washington, DC: Author, 2006. Job performance ratings have often been found to have reliabilities below .70. See Viswesvaran, Chockalingam, Deniz S. Ones, and Frank L. Schmidt. (1996). Comparative Analysis of the Reliability of Job Performance Ratings. *Journal of Applied Psychology,* vol. 88, no. 5, 2007: pp. 557–572.

[11] The estimated reliability of the difference between two different dimension scores, averaged across the four observations in the second year of the performance feedback study, was .35 to .43 and .18 to .30 for the two rubrics used.

For more information on the full studies cited, please visit:

https://ies.ed.gov/ncee/projects/evaluation/tq_performance.asp

and

https://ies.ed.gov/ncee/projects/evaluation/tq_incentive.asp

This brief was prepared for NCEE by Andrew Wayne and Michael Garet of American Institutes for Research and Alison Wellington and Hanley Chiang of Mathematica Policy Research. The work was completed under contracts ED-IES-11-C-0066 (Project Officer: Melanie Ali) and ED-IES-14-C-0115 (Project Officer: Elizabeth Warner).