# Evaluation of the Teacher Incentive Fund: Implementation and Impacts of Pay-for-Performance After Two Years

September 2015

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

U.S. Department of Education

THIS PAGE IS INTENTIONALLY BLANK

# Evaluation of the Teacher Incentive Fund: Implementation and Impacts of Pay-for-Performance After Two Years

**September 2015**

Hanley Chiang
Alison Wellington
Kristin Hallgren
Cecilia Speroni
Mariesa Herrmann
Steven Glazerman
Jill Constantine
**Mathematica Policy Research**

**Elizabeth Warner**
*Project Officer*
Institute of Education Sciences

NCEE 2015-4020
U.S. DEPARTMENT OF EDUCATION

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

**U.S. Department of Education**
Arne Duncan
*Secretary*

**Institute of Education Sciences**
Ruth Neild
*Deputy Director for Policy and Research*
*Delegated Duties of the Director*

**National Center for Education Evaluation and Regional Assistance**
Joy Lesnick
*Acting Commissioner*

**September 2015**

The report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0112-0012. The project officer is Elizabeth Warner in the National Center for Education Evaluation and Regional Assistance.



Chiang, Hanley, Alison Wellington, Kristin Hallgren, Cecilia Speroni, Mariesa Herrmann, Steven Glazerman, Jill Constantine. (2015). *Evaluation of the Teacher Incentive Fund: Implementation and Impacts of Pay-for-Performance After Two Years* (NCEE 2015-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

This report is available on the IES website at http://ies.ed.gov/ncee.

## ACKNOWLEDGMENTS

This study would not have been possible without the contributions of many individuals. We are grateful for the cooperation of many TIF administrators, teachers, principals, district leaders, and central office staff who assisted with the study's data collection and provided important information that shaped the study. A dedicated technical assistance team helped TIF districts implement the programs examined in this study. This team was led by Duncan Chaplin and Jeffrey Max and included Lauren Akers, Kevin Booker, Julie Bruch, Albert Liu, Allison McKie, Debbie Reed, Alex Resch, Christine Ross, and Margaret Sullivan from Mathematica and Patrick Schuermann and Eric Hilgendorf from the Peabody College of Education at Vanderbilt University.

Several individuals made enormous efforts to collect data successfully for this study. Sheila Heaviside and Annette Luyegu provided excellent leadership over our administration of teacher, principal, and district surveys, and Kathy Shepperson oversaw the design of key systems for collecting this survey data. Lauren Akers, Margaret Sullivan, and Claire Smither Wulsin patiently conducted and summarized numerous interviews with TIF administrators. Acquiring and processing administrative data required a large effort led by Jacqueline Agufa with assistance from Michael Brannan, Dylan Ellis, Chris Jones, Serge Lukashanets, Jeremy Page, Nina Punukollu, and Juha Sohlberg.

Many people contributed to the analysis and interpretation of the study's data and the production of this report. The study received useful advice from our technical working group, consisting of David Heistad, James Kemple, Daniel McCaffrey, Anthony Milanowski, Richard Murnane, Jeffrey Smith, and Jacob Vigdor. At Mathematica, Chris Jones helped with a variety of critical tasks ranging from communicating with TIF districts to facilitating the management of the project. The analysis was made possible by an excellent team of programmers, consisting of Raúl Torres Aragon, Michael Brannan, Molly Crofton, John Hotchkiss, and William Leith, with expert guidance from Mary Grider. Cindy George and John Kennedy oversaw the editing of the report, and Jill Miller carefully and patiently prepared the report for publication.

THIS PAGE IS INTENTIONALLY BLANK

## CONTENTS

THIS PAGE IS INTENTIONALLY BLANK

## TABLES

## FIGURES

THIS PAGE IS INTENTIONALLY BLANK

# EXECUTIVE SUMMARY

Recent efforts to attract and retain effective educators and to improve teaching practices have focused on reforming evaluation and compensation systems for teachers and principals. In 2006, Congress established the Teacher Incentive Fund (TIF), which provides grants to support performance-based compensation systems for teachers and principals in high-need schools. This study focuses on performance-based compensation systems that were established under TIF grants awarded in 2010. It examines grantees' programs and implementation experiences and the impacts of pay-for-performance bonuses on educator effectiveness and student achievement.

This report, the second from the study, describes the programs and implementation experiences of all 2010 TIF grantees in the 2012–2013 school year, the second of four years of implementation for nearly all grantees. The main findings for all districts that received 2010 TIF grants include the following:

- **Full implementation of TIF continues to be a challenge, although districts' implementation from the first to the second year improved somewhat.** Although 90 percent of all TIF districts in 2012–2013 reported implementing at least 3 of the 4 required components for teachers, only about one-half (52 percent) reported implementing all four. This was a slight improvement from the first year of implementation, when 85 percent of districts reported implementing at least 3 of the 4 required components and 46 percent reported implementing them all.
- **Near the end of the second year of implementation, most districts reported that sustainability of their TIF program was a major challenge; however, few reported other key activities related to their program were a major challenge.** By the end of 2012–2013, 65 percent of TIF districts reported that sustainability of the program was a major challenge. In contrast, no more than one-third of districts reported that other activities related to their program (such as incorporating student achievement growth into teacher evaluations or conducting observations) were a major challenge.

This report also provides detailed findings from a subset of 2010 TIF grantees, the evaluation districts, that participated in a random assignment study of the pay-for-performance component of TIF. For the ten evaluation districts that completed two years of TIF implementation, the report provides an in-depth analysis of TIF implementation and the impacts of pay-for-performance bonuses on educator and student outcomes after the first (2011–2012) and second (2012–2013) years. The main findings for the ten evaluation districts include the following:

- **Few evaluation districts structured pay-for-performance bonuses to align well with TIF grant guidance.** The grant notice provided guidance, although not specific requirements, about how to structure pay-for-performance bonuses to be substantial, differentiated, and challenging to earn. At least half of the evaluation districts (70 percent in Year 1 and 50 percent in Year 2) met the guidance for awarding differentiated performance bonuses for teachers. However, in each year, no more than 30 percent of districts awarded bonuses for teachers that were substantial or challenging to earn. Likewise, no more than 30 percent of districts awarded bonuses for principals that were differentiated, substantial, or challenging to earn.

- **Educators' understanding of key program components improved from the first to the second year, but many teachers still misunderstood whether they were eligible for performance bonuses or the amount they could earn.** Teachers had a better understanding of how their performance was evaluated in the second year than in the first. For example, about 85 percent of teachers reported being evaluated on at least two classroom observations in the second year compared to about 75 percent of teachers in the first year. In schools that offered pay-for-performance bonuses, teachers' and principals' understanding of their eligibility for bonuses also improved (by 13 and 35 percentage points, respectively). However, many teachers in these schools (38 percent in the second year) still did not understand that they were eligible for a bonus. They also continued to underestimate how much they could earn from performance bonuses, reporting a maximum bonus that was only two-fifths the size of the actual maximum bonuses awarded.
- **Pay-for-performance had small, positive impacts on students' reading achievement; impacts on students' math achievement were not significant but similar in magnitude.** After two years of TIF implementation, the average reading score was 1 percentile point higher in schools that offered pay-for-performance bonuses than in schools that did not. This difference was equivalent to a gain of about three additional weeks of learning.

## TIF Grants and Requirements

From 2006 to 2012, the U.S. Department of Education awarded about $1.8 billion to support 131 TIF grants. Sixteen grants were awarded in 2006, 18 in 2007, 62 in 2010, and 35 in 2012.

The 2010 TIF grants differed from prior TIF grants by providing more detailed guidance on the measures used to evaluate educators and on the design of the pay-for-performance bonuses. The 2010 grants required performance-based compensation systems implemented in districts to include four components. This study focuses most heavily on examining the implementation and impacts of one of those requirements: pay-for-performance bonuses.

### Required Program Components of the Performance-Based Compensation Systems

The four required TIF components are:

1. **Measures of educator effectiveness.** Grantees were required to measure the effectiveness of teachers and principals using students' achievement growth and at least two observations of classroom or school practices. They had discretion to include additional measures.
2. **Pay-for-performance bonus.** Grantees had to offer bonuses to educators based on how they performed on the effectiveness measures. The bonuses aimed to incentivize educators and reward them for being effective in their classrooms and schools. Bonuses had to be substantial, differentiated, challenging to earn, and based solely on educators' effectiveness.
3. **Additional pay opportunities.** The performance-based compensation system had to include pay opportunities for educators to take on additional roles or responsibilities.

These roles might include becoming a master or mentor teacher who directly counsels other teachers or develops or leads professional development sessions for teachers.

4. **Professional development.** TIF grantees were required to support teachers and principals in their performance improvement efforts. Support included providing information about measures on which educators would be evaluated and more targeted professional development based on an educator's actual performance on the effectiveness measures.

### The TIF Grant Competition

The 2010 TIF grant notice differed from the other rounds in that it included a main and an evaluation competition (Max et al. 2014). By holding two separate competitions, the U.S. Department of Education identified a group of grantees that, by virtue of having applied for an evaluation grant, had indicated their interest and willingness to participate in a more in-depth evaluation of their TIF grants.

A key difference between the non-evaluation and evaluation grantees is that applicants for the evaluation grants received more specific guidance about the structure of their pay-for-performance bonuses. They received examples of pay-for-performance bonuses that were *substantial* (with an average bonus worth 5 percent of the average educator's salary), *differentiated* (with at least some educators expecting to receive a bonus worth three times the average payout), and *challenging* to earn (with only those performing significantly better than average receiving bonuses). Although applicants had discretion over the proposed structure of the pay-for-performance bonus, these examples provided additional guidance to evaluation grant applicants and might have influenced how they designed their performance-based compensation systems.

Applicants for evaluation grants had to meet the same requirements for the performance-based compensation system as non-evaluation grantees and some additional requirements. One important requirement was that evaluation grant applicants had to agree to participate in a random assignment evaluation of pay-for-performance bonuses. Schools within a district were randomly assigned to implement either all four required components of the performance-based compensation system, including pay-for-performance bonuses (the treatment group), or all components *except* pay-for-performance bonuses (the control group).

## The TIF Study

The purpose of this multiyear study is to describe the program characteristics and implementation experiences of 2010 TIF grantees and estimate the impact of pay-for-performance bonuses within a well-implemented, performance-based compensation system. Because educators' understanding of and responses to this policy can change over time, this study plans to follow the grantees for the full duration of the five-year grants.

The study is addressing four research questions:

1. What are the characteristics of all TIF districts and their performance-based compensation systems? What implementation experiences and challenges did TIF districts encounter?
2. How do teachers and principals in schools that did or did not offer pay-for-performance bonuses compare on key dimensions, including their understanding of TIF program

features, exposure to TIF activities, allocation of time, and attitudes toward teaching and the TIF program?

3. How do pay-for-performance bonuses affect educator effectiveness and the retention and recruitment of high-performing educators?
4. What is the impact of pay-for-performance bonuses on students' achievement on state assessments in math and reading?

This report is the second of four planned reports from the study. The first report (Max et al. 2014) addressed the first two research questions based on information from the 2011–2012 school year. This second report uses information from the first (2011–2012) and second (2012–2013) years of TIF implementation to describe the ways in which evaluation districts structured the components of their programs and communicated information about those components (question 1). This report also captures the views, attitudes, and behaviors of educators as they evolved over two years of implementation (question 2) and presents initial impacts of pay-for-performance on educator effectiveness and student achievement after the first and second years (questions 3 and 4).

**Districts in the Study**

Although this report provides the greatest amount of information on the evaluation districts, it also provides a broad overview of TIF implementation by all 2010 grantees in the 2012–2013 school year. This analysis was based on 155 districts that participated in TIF in 2012–2013.

This report's in-depth analyses of TIF implementation and the effects of pay-for-performance on educator and student outcomes were based on information from the evaluation districts. Of the 13 evaluation districts, 10 completed two years of TIF implementation—2011–2012 and 2012–2013—during the period covered by the report. The remaining 3 evaluation districts completed their first year of TIF implementation in 2012–2013. This report focuses primarily on the 10 districts for which data were available on two years of TIF implementation. Focusing on districts that completed two years of TIF implementation enabled us to examine changes in educators' perceptions and practices from the first to the second year and assess whether impacts on educator and student outcomes also evolved during that time.

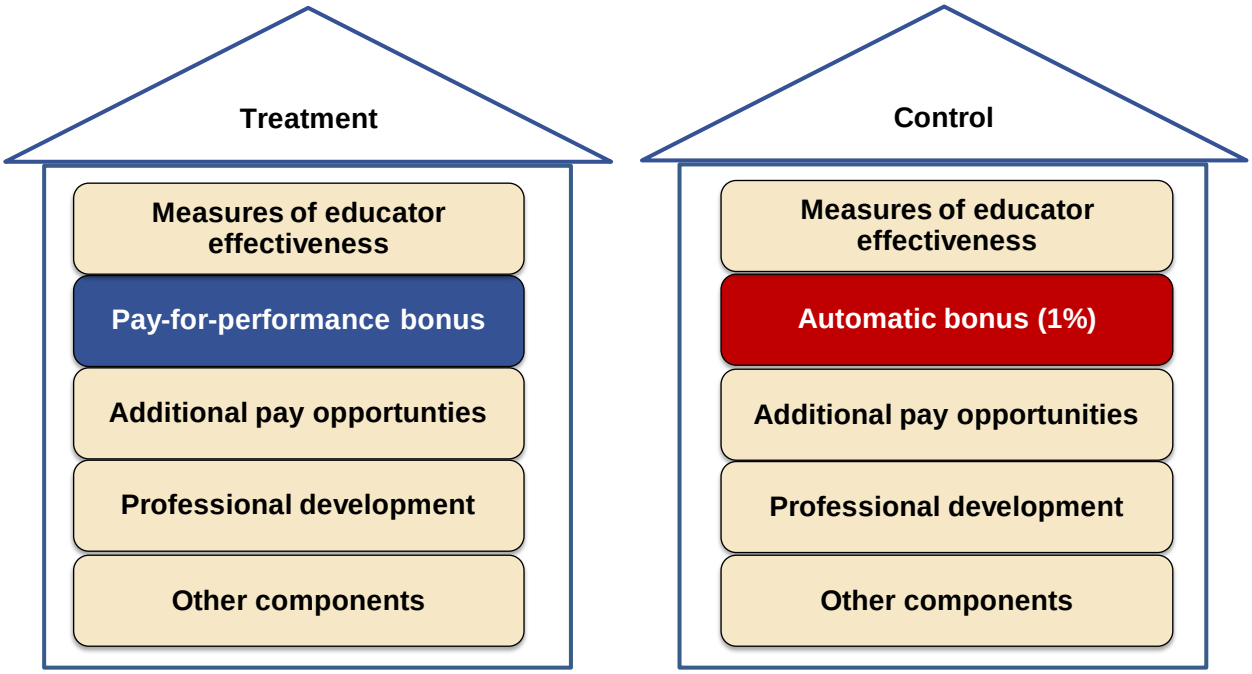
**Experimental Study Design**

The study used an experimental study design to assess the impacts of pay-for-performance on educator and student outcomes. Elementary and middle schools within the evaluation districts were assigned randomly—that is, completely by chance—to treatment and control groups. As shown in Figure ES.1, treatment and control schools were expected to implement the same required components of the district's performance-based compensation system, except for the pay-for-performance bonus component. As a result, the study measured the impact of pay-for-performance bonuses implemented within the context of broader performance-based compensation systems. The study was not designed to measure the impact of implementing a TIF grant or the multiple components of a performance-based compensation system.

Teachers and principals in treatment schools were eligible to earn a pay-for-performance bonus; teachers and principals in control schools received an automatic bonus worth approximately 1 percent of their annual salary. The 1 percent bonus ensured that all educators in evaluation schools received some benefit from participating in the study: either the opportunity to earn a pay-for-performance

bonus or the automatic bonus. Therefore, the impact of pay-for-performance estimated in this study potentially reflected two key differences between treatment and control schools: (1) bonuses in treatment schools were differentiated based on performance; and (2) bonuses in treatment schools were larger, on average, than in control schools.

**Figure ES.1. Random Assignment Evaluation Design**




The key advantage of this study's random assignment design is that, at the beginning of the study, the treatment and control groups were expected to include students and educators with similar characteristics. Because the two groups were expected to differ only in the opportunity for educators to receive pay-for-performance bonuses, differences in outcomes between the groups could be attributed to the impact of pay-for-performance.

## Schools in the Study

Analyses of educator and student outcomes were based on 132 schools—66 treatment schools and 66 control schools—that implemented the TIF program for two years. Before random assignment, evaluation districts chose which schools to include in the evaluation. Because a primary objective of the study was to measure the impact of pay-for-performance on student achievement on state assessments in high-need schools, every participating school had to have (1) at least half of its students receiving free or reduced-price lunch and (2) at least one grade level tested by state assessments (3rd to 8th grade).

## Data Sources

Data for this report came from multiple sources. The sources enabled us to examine implementation broadly in all TIF districts and, within evaluation districts, to report on more detailed aspects of implementation and the impacts of pay-for-performance on educator and student outcomes.

**Data on all 2010 TIF districts.** The study team collected data on all TIF districts from two sources. First, to compare characteristics of evaluation and non-evaluation districts, the study team used information from the Common Core of Data. Second, to describe broadly the TIF program features that districts reported implementing and the challenges they encountered in implementation, the study team administered a survey to all TIF district administrators in 2011–2012 and 2012–2013.

**Additional data on evaluation districts.** We obtained more detail on TIF programs and implementation experiences from interviews with district staff and technical assistance documents. To examine educators' attitudes toward their job and the TIF program, the study team administered surveys to all principals and a sample of teachers in treatment and control schools in spring 2012 and spring 2013. We collected districts' administrative records on teachers and principals to describe their performance ratings, bonuses, and additional pay, as well as to examine the impact of pay-for-performance on educator effectiveness. Finally, to assess the impact of pay-for-performance on student achievement, the study team collected districts' administrative records on students enrolled in treatment and control schools.

## Methods

The study team used several different methods to describe the implementation of TIF and measure the impact of pay-for-performance on educators' and students' outcomes.

**Describing TIF implementation in all 2010 TIF districts.** To describe broadly the program characteristics and implementation challenges reported by all 2010 TIF districts, we summarized their responses to the district survey with means or percentages, as appropriate.

**Describing TIF implementation in evaluation districts.** We conducted a variety of analyses to provide an in-depth description of TIF implementation in the evaluation districts. First, as in the analysis of all 2010 TIF districts, we summarized evaluation districts' survey responses about program characteristics and implementation challenges, but we also supplemented these data with information from telephone interviews and technical assistance documents. Second, to describe educators' actual bonus amounts and performance ratings, we summarized administrative data with means, maximum levels, or percentages of educators receiving particular bonus amounts or ratings. Third, to describe educators' understanding of and experiences with the required TIF components, we summarized educators' survey data, making comparisons between treatment and control schools and across years.

**Measuring the impacts of pay-for-performance on educator and student outcomes.** Within the evaluation districts, we assessed the impacts of pay-for-performance on several educator and student outcomes, including educators' attitudes and behaviors (measured by survey responses), educator effectiveness (measured by performance ratings that educators received from their districts), and student achievement (measured by scores on state assessments in math and reading). For each outcome, we compared the outcomes of educators and students in treatment schools to those of educators and students in control schools. Because the study used random assignment, any differences in educator or student outcomes between the treatment and control groups could be attributed to pay-for-performance and not some other characteristic of the districts or schools.

## Detailed Summary of Findings

### Programs and Experiences of All 2010 TIF Districts

As a comprehensive program for reforming educator compensation and improving educator effectiveness, TIF programs were designed to have multiple, interrelated components. Our analysis of implementation in all 155 TIF districts sought to determine whether they could put into place such a comprehensive system, and whether they faced particular challenges doing so.

**Full implementation of TIF continues to be a challenge, although districts' implementation from the first to the second year improved somewhat.** Although 90 percent of all TIF districts in the second year (2012–2013) reported implementing at least 3 of the 4 required components for teachers, about one-half (52 percent) reported implementing all four. This was a slight improvement from the first year (2011–2012), when 85 percent of districts reported implementing at least 3 of the 4 required components and 46 percent reported implementing them all. More than half of the districts (58 percent in Year 1 and 60 percent in Year 2) implemented all required components for principals aside from professional development, a component for which data were not available.

**Most districts implemented each individual required component of TIF, but were less likely to report offering targeted professional development and evaluating teachers and principals using both student achievement growth and at least two observations**. In Year 2, nearly all the districts (over 90 percent) reported offering teachers and principals bonuses based on their performance and offering educators opportunities to earn additional pay (Table ES.1). In contrast, approximately three-quarters of the districts reported that they offered the required professional development to their teachers and 80 percent reported using both student achievement growth and classroom observations to measure teacher effectiveness (Table ES.1). Fewer districts (65 percent) reported using both student achievement growth and observations of school practices to measure principal effectiveness. Although all districts were expected to evaluate educators using student achievement growth, districts could choose how to measure student achievement growth to evaluate their educators. In Year 2, almost all TIF districts (about 88 percent) reported using an achievement growth measure based on all students in the school to evaluate teacher and principal effectiveness. Fewer districts (64 percent) reported evaluating teachers based on the achievement growth of only the students in their classrooms.

**Near the end of the second year of implementation, most districts reported that sustainability of their TIF program was a major challenge; however, few reported other key activities related to their program were a major challenge.** By the end of 2012–2013, 65 percent of TIF districts reported that sustainability of the program was a major challenge. In contrast, fewer than one-third of districts reported that linking student growth data to teachers (30 percent), explaining student achievement growth to teachers (28 percent), and calculating student achievement growth to evaluate teachers (28 percent) were major challenges. Likewise, only one-third of districts (33 percent) reported that providing useful and timely feedback on student achievement measures was a major challenge.

**Table ES.1. Districts' Reported Implementation of TIF Required Components for Teachers in Year 2 (Percentages)**

| | All 2010 TIF Districts | Evaluation Districts |
|---|---|---|
| **Requirements** | | |
| Requirement 1: Measures of educator effectiveness[a] | 80 | 100 |
| Requirement 2: Pay-for-performance bonus | 98 | 100 |
| Requirement 3: Additional pay opportunities | 91 | 100 |
| Requirement 4: Professional development | 74 | 70 |
| Implemented all requirements | 52 | 70 |
| **Number of Districts—Range[b]** | **142-155** | **10** |

Source: District surveys and district interviews, 2013.

[a]TIF districts were required to use student achievement growth and at least two observations by trained observers to evaluate teachers and principals.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

## TIF Implementation in Evaluation Districts

In-depth information from the evaluation districts enabled the study team to examine, in greater detail, whether the components of their programs provided incentives and supports for educators to improve their effectiveness, and whether educators understood those components.

### *Program Implementation*

As a first step, we examined the extent to which evaluation districts implemented the four required components. We also examined the types of measures that districts used to evaluate educators' effectiveness and described educators' actual performance on those measures, focusing on whether different measures were consistent with each other in assessing how well teachers performed.

**Most evaluation districts reported implementing all required components for teachers. The only component not consistently implemented was professional development.** In Year 2, all evaluation districts reported using measures of effectiveness for teachers and principals that included student achievement growth and at least two observations of classroom or school practices, offering bonuses based on how educators performed on effectiveness measures, and offering additional pay to take on extra roles or responsibilities. Seven of 10 evaluation districts reported providing the required professional development for teachers (Table ES.1).

**All evaluation districts reported using the achievement growth of all students in a school to evaluate teachers, and some also chose to evaluate teachers based on the achievement growth of the students they teach.** Slightly more than half (60 percent) of evaluation districts reported evaluating teachers based on achievement growth in their classrooms. Within these districts, fewer than half (about 40 percent) of teachers were evaluated on the achievement growth of their students.

**Student achievement growth and observation ratings sometimes identified the same educators as high-performing, but many earned higher ratings on observations than on achievement growth.** For example, in Year 2 teachers who scored high (in the top quarter of the rating scale) on achievement growth in their schools were twice as likely to score high on classroom

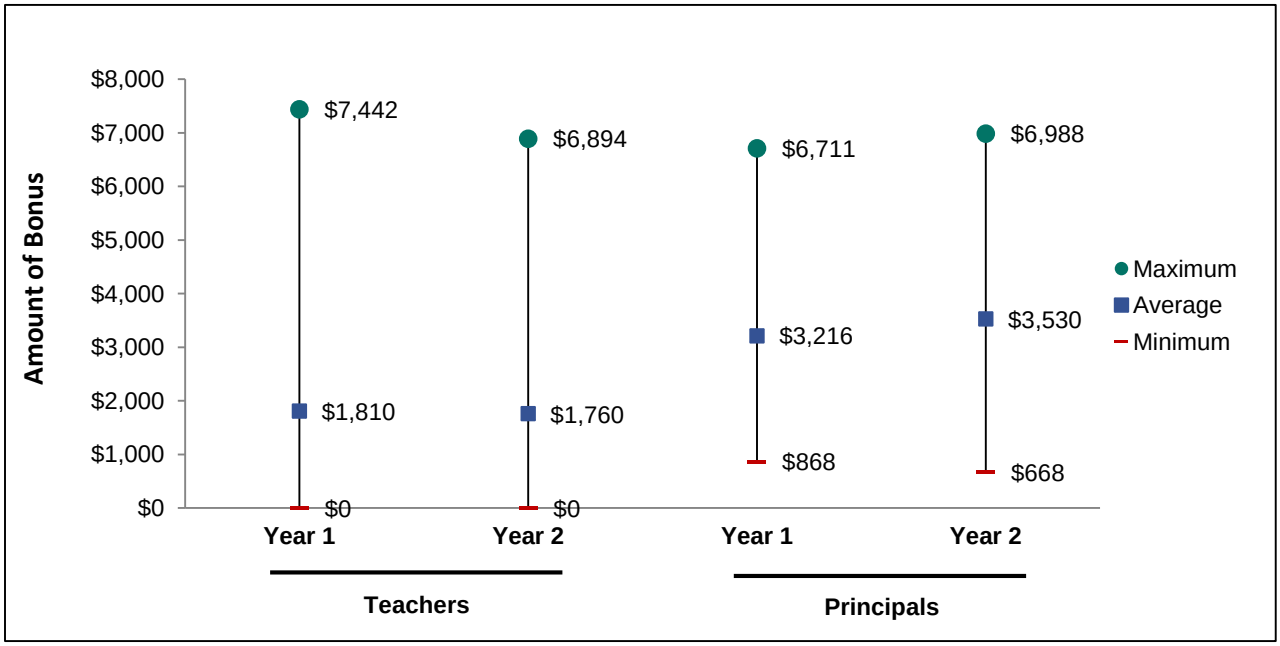
observations compared with teachers who scored low (in the bottom quarter of the rating scale) on achievement growth (38 versus 19 percent). Nevertheless, many teachers (87 percent) who scored low on achievement growth earned at least moderately high ratings on classroom observations, scoring in the top half of the observation rating scale. Likewise, many principals (78 percent) who scored low on achievement growth earned at least moderately high ratings on observations.

### *Pay-for-Performance Bonuses*

The purpose of offering performance bonuses to teachers and principals was to motivate them to improve and reward educators for being effective in their classrooms and schools. To achieve this objective, the TIF notice required that the bonuses had to be substantial, differentiated, and challenging to earn. In this section we examine how well the evaluation districts met this TIF grant guidance.

**At least half of districts met the TIF grant guidance for awarding differentiated pay-for-performance bonuses for teachers, but not the guidance for awarding bonuses that were substantial or challenging to earn.** On average across evaluation districts, the maximum bonus ($7,442 in Year 1 and $6,894 in Year 2) was more than three times the average bonus ($1,810 in Year 1 and $1,760 in Year 2), consistent with the example of a differentiated bonus provided in the TIF grant notice (Figure ES.2). However, the average bonus was about 4 percent of the average teacher's salary—less than the 5 percent guidance for substantial bonuses specified in the TIF grant notice. Fewer than one-third of the districts met the guidance for bonuses that were challenging to earn. Across districts, on average, more than 60 percent of teachers in treatment schools received a bonus in each year.

**Figure ES.2. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers and Principals in Years 1 and 2**




Source: District administrative data (N = 2,189 teachers in Year 1; N = 2,207 teachers in Year 2; N = 65 principals in Year 1; and N = 68 principals in Year 2 in treatment schools).

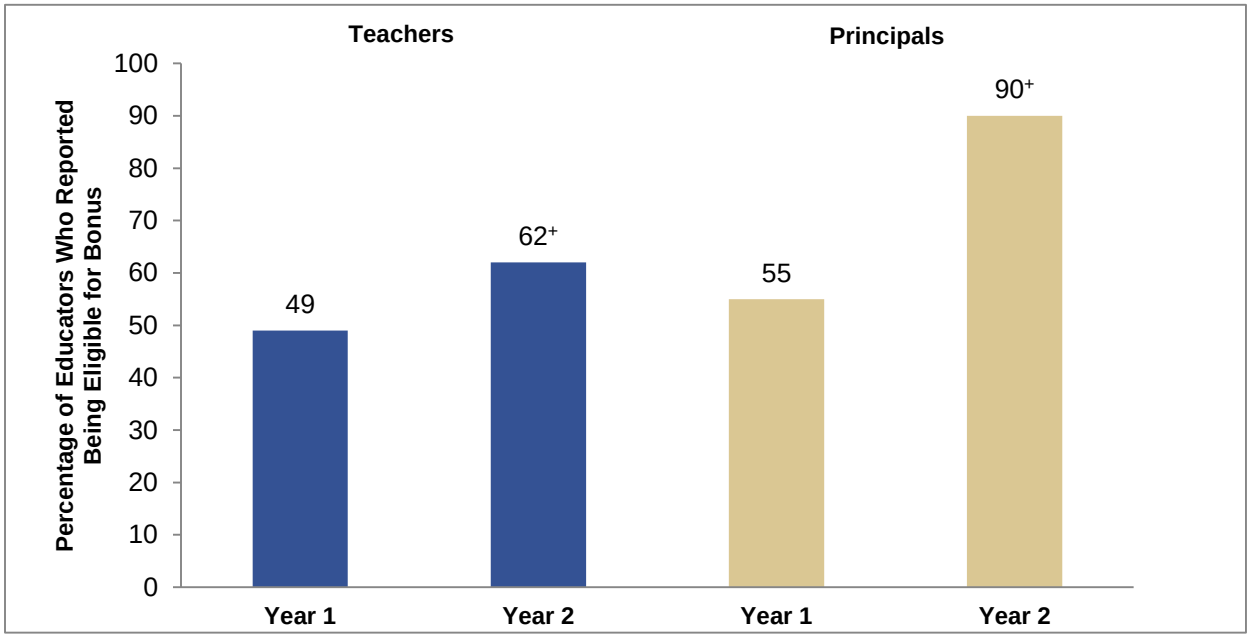
Figure reads: In Year 2, on average across the evaluation districts, the minimum pay-for-performance bonus for teachers was $0, the average pay-for-performance bonus for teachers was $1,760, and the maximum pay-for-performance bonus for teachers was $6,894.

### *Teachers' and Principals' Understanding of Key Components*

In addition to determining how to implement the required components of TIF, districts had to effectively communicate information about those components to educators. Educators' understanding of the components determines how the program can influence educators' behaviors and, ultimately, student achievement.

**Educators' understanding of key program components improved from the first to the second year, but many teachers still misunderstood whether they were eligible for performance bonuses or the amount they could earn.** Teachers had a better understanding of how their performance was evaluated in Year 2 than in Year 1. For example, about 85 percent of teachers reported being evaluated on at least two classroom observations in Year 2, compared to about 75 percent of teachers in Year 1. In treatment schools, teachers' and principals' understanding of their eligibility for bonuses also improved (by 13 and 35 percentage points, respectively; Figure ES.3). However, many teachers in treatment schools (38 percent in Year 2) still did not understand that they were eligible for a bonus. They also continued to underestimate how much they could earn from performance bonuses, reporting a maximum bonus that was only two-fifths the size of the actual maximum bonuses awarded (Figure ES.4). Principals also continued to underestimate the potential amount of performance bonuses they could receive, but their expectations were better aligned with actual bonus amounts than were teachers' expectations. In Year 2, principals in treatment schools, on average, reported that the maximum pay-for-performance bonus they could receive was 87 percent of the actual maximum bonus awarded to principals.

**Figure ES.3. Teachers and Principals in Treatment Schools Who Reported Being Eligible for Pay-for-Performance Bonuses (Percentages)**

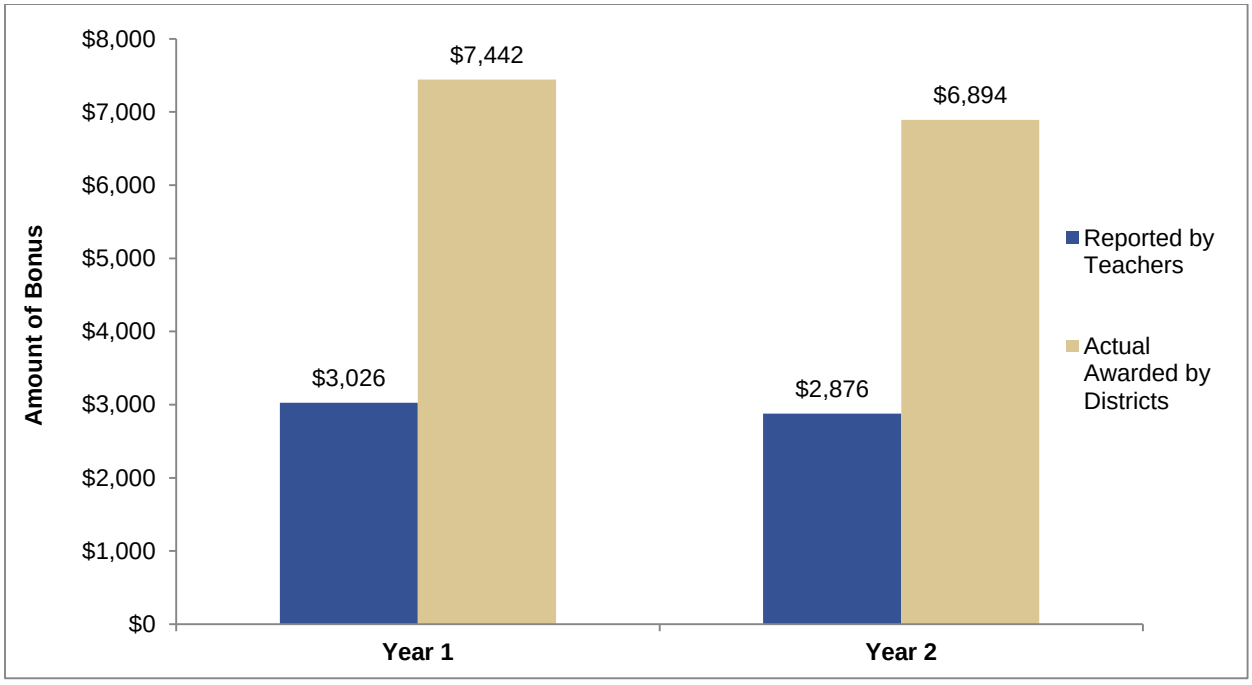



Source: Teacher and principal surveys, 2012 and 2013 (N = 377 teachers in Year 1; N = 444 teachers in Year 2; N = 64 principals in Year 1; and N = 63 principals in Year 2).

Figure reads: Among teachers in treatment schools, 49 and 62 percent reported being eligible for a pay-for-performance bonus in Year 1 and Year 2, respectively.

+Difference between 2011–2012 and 2012–2013 is statistically significant at the .05 level, two-tailed test.

**Figure ES.4. Reported and Actual Maximum Pay-for-Performance Bonuses for Teachers in Treatment Schools in Years 1 and 2**




Source: Teacher surveys (2012 and 2013) and educator administrative data (N = 223 teachers in Year 1; N = 232 teachers in Year 2; N = 10 districts).

Figure reads: In Year 2, teachers in treatment schools reported, on average, that the maximum pay-for-performance bonus they could earn was $2,876. On average across districts, the actual maximum bonus districts awarded to teachers in Year 2 was $6,894.

## Impacts of Pay-for-Performance on Educators' Attitudes and Behaviors

The ways in which pay-for-performance programs affect educators' attitudes (such as job satisfaction) and behaviors (such as allocation of time) can shape how pay-for-performance affects student outcomes. For example, pay-for-performance could motivate educators to improve their effectiveness if it makes them more satisfied with pay opportunities and the feedback they receive on performance evaluations. However, if the presence of pay-for-performance discourages useful collaboration, lowers morale, or makes a school less appealing to effective educators, it could have a negative effect on the work environment and, ultimately, on student achievement.

**Most teachers and principals reported being satisfied with their professional opportunities, how they were evaluated, and their school environment.** For example, in Year 2, at least 80 percent of teachers reported being satisfied with their opportunities to enhance their skills, their quality of interaction with colleagues, and colleagues' efforts. The percentage of principals satisfied with aspects of their professional opportunities, evaluation system, and school environment ranged from 61 to 90 percent.

**Educators in treatment schools tended to be less satisfied than educators in control schools, with one exception; teachers in treatment schools were more satisfied with their opportunities to earn extra pay.** For example, in Year 2, a lower percentage of teachers in treatment schools than control schools were satisfied with the use of student achievement scores to assess their performance (60 versus 69 percent) and with the feedback received on their performance (75 versus

80 percent). In Year 2, principals in treatment schools were less satisfied than principals in control schools with the use of observations to assess their skills (61 versus 85 percent) and the use of student achievement scores to assess performance (66 versus 82 percent.) The one exception to the pattern of lower satisfaction in treatment schools was that more treatment teachers were satisfied with their opportunities to earn extra pay (62 versus 54 percent) in Year 2.

**Most teachers had positive attitudes toward the TIF program, but teachers in treatment schools were less likely than teachers in control schools to be positive about TIF.** In both years of TIF implementation, about two-thirds of teachers were glad they were participating in TIF and at least half felt TIF was fair. However, treatment teachers in Year 2 were more likely than control teachers to report that TIF reduced their freedom to teach the way they would like (40 versus 30 percent), harmed the collaborative nature of teaching (29 versus 21 percent), and caused increased pressure to perform (65 versus 51 percent).

**Impacts of Pay-for-Performance on Educator Effectiveness and Student Achievement**

A central objective of the TIF grants is to improve student achievement in high-need schools by increasing educator effectiveness—in particular, by enabling schools to attract and retain more effective educators and motivating educators to improve their effectiveness. This study measured educator effectiveness using the performance ratings that educators received from their districts, and measured student achievement using students' reading and math scores on state assessments.[1]

**Pay-for-performance led to teachers and principals earning higher effectiveness ratings based on student achievement growth in their schools, but did not affect ratings based on observations of their classroom or school practices.** The school achievement growth ratings of teachers and principals in treatment schools were 0.34 points higher than those of teachers and principals in control schools (based on a scale ranging from 1 to 4) in Year 1, and 0.25 points higher in Year 2. In Years 1 and 2, treatment and control teachers earned similar classroom observation ratings, and treatment and control principals earned similar ratings from observations of their school practices.

**Pay-for-performance did not enable schools to retain or attract more higher-performing teachers.** Teachers who stayed at treatment and control schools over the first two years of TIF implementation were similar in effectiveness, as measured by their classroom observation ratings and classroom achievement growth ratings. Teachers newly hired at treatment and control schools also earned similar ratings.

**Pay-for-performance led to more higher-performing principals staying in their schools and more lower-performing principals leaving their schools.** School achievement growth ratings were higher among principals who stayed at treatment schools than those who stayed at control schools over the first two years of TIF implementation. Observation ratings were lower among principals who left treatment schools than those who left control schools.

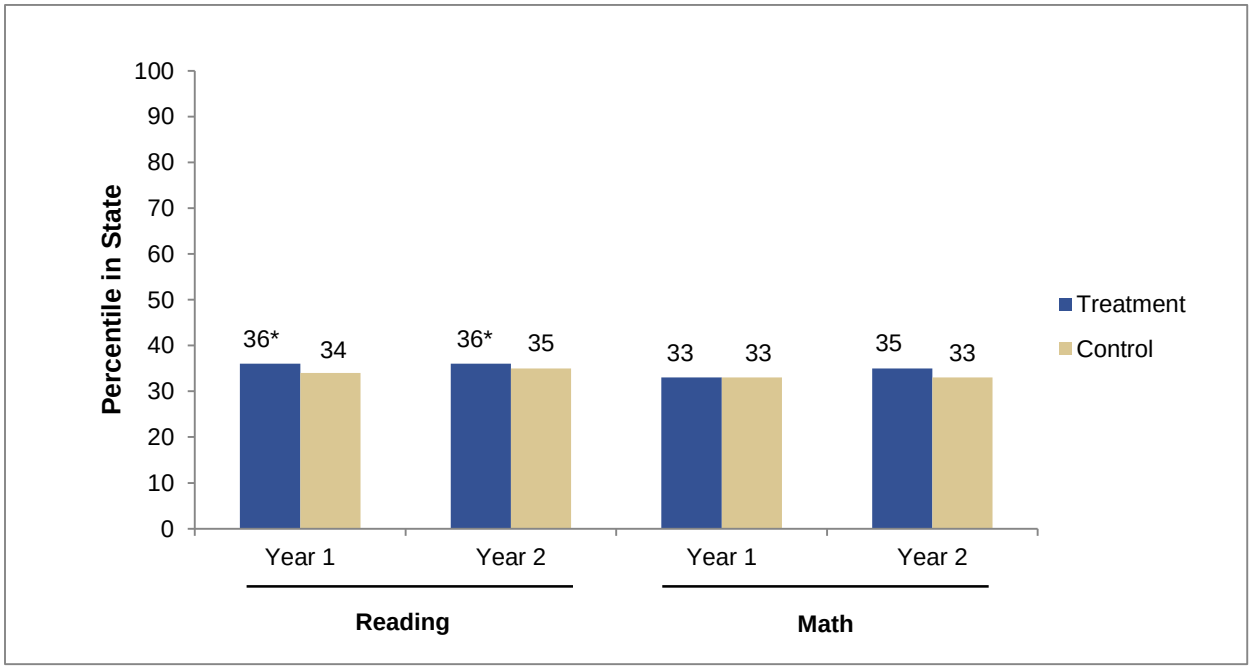
**Pay-for-performance had small, positive impacts on students' reading achievement; impacts on students' math achievement were insignificant but similar in magnitude.** In Years 1 and 2, the average student in a treatment school earned a reading score at approximately the 36th

[1] This study examined the impacts of pay-for-performance bonuses on the average outcomes of schools that offered those bonuses, but for simplicity we describe these findings as impacts on educators' or students' outcomes.

percentile in his or her state, whereas the average student in a control school scored at approximately the 34th or 35th percentile—a difference of 1 to 2 percentile points (Figure ES.5). This difference translated to a gain of about 3 weeks of additional learning in a typical 36-week school year. In math, differences in student achievement between treatment and control schools after Year 2 were not statistically significant, but were also positive and similar in magnitude to those in reading.

**The impacts of pay-for-performance on student achievement differed among districts, but differences in impacts were not related to differences in key program characteristics measured by this study.** The impacts of pay-for-performance on reading and math achievement were not related to a variety of program and implementation characteristics, including (1) the use of student achievement growth in teachers' own classrooms to measure teacher effectiveness, (2) teachers' understanding of their eligibility for performance bonuses, and (3) the timing of bonus notification and award.

**Figure ES.5. Average Student Achievement in Treatment and Control Schools After Years 1 and 2 (Percentiles)**




Source: Student administrative data (N = 40,576 students for Year 1 reading; N = 40,391 students for Year 2 reading; N = 40,852 students for Year 1 math; and N = 40,709 students for Year 2 math).

Figure reads: In Year 2, students in treatment schools earned an average reading score at the 36th percentile in their state, and students in control schools earned an average reading score at the 35th percentile.

*Difference between treatment and control schools is statistically significant at the .05 level, two-tailed test.

## Concluding Thoughts

Overall, the 2010 TIF districts were able to implement most required components of a comprehensive performance-based compensation system without major, widespread challenges. However, many districts still did not put into place all the required components by the end of the second year of implementation. In addition, TIF districts were expected to sustain their programs beyond the life of the grant, but, midway through their grant, many TIF districts already reported that sustaining their programs would be a major challenge.

A primary objective of TIF grants is to raise student achievement in high-need schools. Based on the experiences of ten districts that participated in the national evaluation and completed two years of program implementation, the pay-for-performance component of TIF made a small contribution toward achieving this objective. Pay-for-performance bonuses generated slightly higher student reading achievement, and gains in math were similar in magnitude but not statistically significant.

The driving principle behind TIF is that increasing educator effectiveness is the key to raising student achievement and pay-for-performance bonuses are one way to increase educator effectiveness. We confirmed that the positive impact of pay-for-performance on student achievement was also reflected in positive impacts on educator effectiveness, as measured by the effectiveness ratings that educators received from their districts. Increases in educator effectiveness could have occurred either because teachers and principals improved their own effectiveness or because staffing changes resulted in more effective educators choosing to work at schools with pay-for-performance. We found little evidence for changes in staffing among teachers. Among principals, we found some evidence that pay-for-performance caused more high performers to stay at their schools and more low performers to leave their schools after the first year of TIF implementation. However, it is unclear whether these staffing changes among principals actually contributed to the positive impacts of pay-for-performance on student achievement. The positive impacts on reading achievement materialized in the first year—before principals had the opportunity to remain at or leave their schools—and the impacts did not increase from the first to the second year. The remaining explanation for why pay-for-performance raised student achievement in the first two years of TIF implementation is that it caused educators to improve their performance.

Many factors could have contributed to the size of the impact of pay-for-performance bonuses on student achievement. For example, the theory behind awarding educators performance bonuses to improve student achievement assumes that the prospect of earning a bonus can motivate educators to change their practices. Although pay-for-performance made teachers more satisfied with their opportunities to earn extra pay, it made teachers less satisfied with factors associated with how they were evaluated, their school environment, and their TIF program. These effects would have had offsetting impacts in shaping educators' motivation to work more effectively or to work in schools that offer performance bonuses. Furthermore, implementation findings indicated that many educators were unaware of important aspects of their program. For example, educators' understanding of program components improved between the first and second years, yet many teachers in treatment schools still did not understand that they were eligible for a performance bonus or underestimated how much they could earn from these bonuses. It is also unclear whether the actual structure of the bonuses would have provided educators with an incentive to modify their classroom or school practices, given that most educators received a bonus and average bonuses were not large.

Evidence from future years will provide more clarity on whether, over a longer period, the impacts of pay-for-performance evolve as educators continue accumulating more understanding of and experience with this program.

# I. INTRODUCTION

Recent efforts to attract and retain effective educators and to improve teacher practices have focused on reforming evaluation and compensation systems for teachers and principals. In 2006, Congress established the Teacher Incentive Fund (TIF), which provides grants to support performance-based compensation systems for teachers and principals in high-need schools. The TIF grants have two goals:

- Reform compensation systems to reward educators for improving student achievement
- Increase the number of high-performing teachers in high-need schools and hard-to-staff subject areas

The incentives and support offered through TIF grants aim to improve student achievement by improving educator effectiveness and the quality of the teacher workforce.

This is the second of four planned reports from a multiyear study focusing on the TIF grants awarded in 2010.[1] The first report (Max et al. 2014) examined grantees' implementation experiences and intermediate educator outcomes near the end of the first year of program implementation, before the first pay-for-performance bonuses were awarded to teachers and principals. This second report examines grantees' implementation experiences and educators' understanding of, and attitudes toward, the program near the end of the second year of program implementation, as well as changes in educators' understanding and attitudes. This report also examines the impacts of pay-for-performance bonuses on educator effectiveness and student achievement after one and two years of TIF implementation.

This study has two main goals. First, it will inform program development and improvement by describing how grantees implemented their performance-based compensation systems and the implementation challenges they faced. Second, it will test whether pay-for-performance bonuses lead to increases in educator effectiveness and student achievement.

## Previous Research on Pay-for-Performance Programs for Educators

Research on the effectiveness of pay-for-performance initiatives is inconclusive. Few studies of U.S. pay-for-performance programs have found consistent impacts on student achievement, and fewer still have examined the impact of pay-for-performance bonuses on teacher retention and recruitment (Max et al. 2014).[2]

[1] TIF grants often are referred to by the round of the grant award. TIF 1, TIF 2, TIF 3, and TIF 4 correspond to the 2006, 2007, 2010, and 2012 grant awards, respectively. For this report, all references to TIF are for the 2010 awardees.

[2] Max et al. (2014) includes a summary of previous research. Since that report was written, several nonexperimental studies have been published that examine the association between a comprehensive financial incentive program that includes pay-for-performance, and student achievement. Two studies examined the association between Denver's TIF-funded ProComp program and student achievement and teacher mobility (Goldhaber and Walch 2012; Fulbeck 2014). Goldhaber and Walch (2012) found mixed evidence that participating in the ProComp program was associated with improved student achievement. Fulbeck (2014) found that receiving a financial incentive through ProComp decreased the likelihood that a teacher left the district. Sojourner et al. (2014) examined the association between Minnesota's Q-Comp program and student achievement. The authors found a positive association between Q-Comp and students' reading

Although evidence is growing, there still are few high quality studies of comprehensive, well-implemented pay-for-performance programs. Therefore, many unanswered questions remain about the possible effects of pay-for-performance programs similar to those designed and supported by TIF grants. Areas of concern of previous studies include the following:

- **Study design limitations.** Many studies used a nonexperimental design, meaning they did not rely on random assignment (Dee and Wyckoff 2013; Goldhaber and Walch 2012; Fulbeck 2014; Sojourner et al. 2014; Springer et al. 2014; Springer et al. 2009a, 2009b; Slotnik et al. 2013; Shifrer et al. 2013; Bayonas 2010). These studies leave open the possibility that observed outcomes are due to unobserved school, educator, or student characteristics, rather than the offer of pay-for-performance programs. All the experimental studies included schools from only one school district, making it difficult for policymakers to determine whether the study findings can be generalized more broadly (Marsh et al. 2011; Fryer 2011; Goodman and Turner 2011; Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012; Springer et al. 2010; Springer et al. 2012; Fryer et al. 2012).

- **Potential design weaknesses of pay-for-performance programs.** One or more design weaknesses existed in some of the pay-for-performance programs previously studied. For example, the average and maximum pay-for-performance bonuses may have been too small to provide meaningful incentives for teachers to change their practices (Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012; Springer et al. 2009a, 2009b). In some cases, teachers received similar bonuses regardless of their measured effectiveness (Marsh et al. 2011; Fryer 2011; Goodman and Turner 2011; Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012). Finally, some programs awarded bonuses to a high percentage of eligible teachers, perhaps diminishing their motivation to alter their teaching practices (Marsh et al. 2011; Fryer 2011; Goodman and Turner 2011; Shifrer et al. 2013). In addition, communication about the program was, in some cases, very limited (Springer et al. 2010), or the program itself was complicated to explain (Goldhaber and Walch 2012; Fulbeck 2014).

- **Pay-for-performance programs varied on the inclusion of other design features that may influence educator and student outcomes.** Pay-for-performance bonuses may work to improve student achievement only if they are part of a more comprehensive reform package that helps teachers effectively change their teaching practices. Some of the studies examined the impact of pay-for-performance within the context of these more comprehensive reforms (Goldhaber and Walch 2012; Fulbeck 2014; Sojourner et al. 2014; Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012; Bayonas 2010; Slotnik et al. 2013; Springer et al. 2014); others did not (Fryer et al. 2012; Springer et al. 2010; Marsh et al. 2011; Fryer 2011; Goodman and Turner 2011). Similarly, the criteria for earning pay-for-performance bonuses may affect the impact that bonuses have on teacher practices. For example, pay-for-performance bonuses based only on a teacher's ability to raise his or her own students' test scores may not encourage collaboration or may negatively affect school morale. On the other hand, pay-for-performance bonuses that rely on student

achievement, but mixed evidence on the association between the program and students' math achievement. Balch and Springer (2015) examined the association between Austin, Texas' REACH program and students' test score gains in math and reading. They found that REACH was positively associated with test score gains in math and reading in the first year of implementation and these gains were maintained (but did not grow) in the second year of implementation.

achievement growth within an entire school may discourage individual teachers from changing their behaviors. Only two of the programs that were evaluated using an experimental study included group- and school-based incentives as well as individual teacher incentives (Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012; Fryer et al. 2012).

Previous research on the design, implementation, and effects of pay-for-performance has informed the design and evaluation of the TIF grants. In addition, targeted technical assistance supported program implementation to help ensure programs were well designed. This report will be the first study to present findings from a large, multisite random assignment study of the impact of pay-for-performance, as part of a comprehensive reform system, on educator effectiveness and student achievement.

In the following sections, we provide a framework for the evaluation by describing key components of TIF grants and presenting a logic model of how pay-for-performance could influence student outcomes.

## TIF Grant Competition

From 2006 to 2012, the U.S. Department of Education (ED) awarded about $1.8 billion to support 131 TIF grants. ED awarded 16 grants in 2006, 18 in 2007, 62 in 2010, and 35 in 2012. The TIF grants awarded in 2010 ranged from $607,211 to $62,325,746 over a five-year period.[3] Among the 62 TIF grantees in 2010, more than two-thirds were states or school districts (69 percent), 16 percent were nonprofits, 13 percent were charter schools or charter management organizations, and 2 percent were universities. Grantees that were not states or school districts had to partner with a state or local education agency. The 2010 grants were supported, in part, by the American Recovery and Reinvestment Act of 2009 (ARRA). As part of this funding, Congress required a rigorous evaluation of the 2010 grantees, which are the focus of this report.

The 2010 TIF grants were designed to create comprehensive performance-based compensation systems that could provide (1) incentives for educators to become more effective in improving student achievement in high-need schools, and (2) support for educators to improve their performance. The 2010 TIF grants differed from prior TIF grants by providing more detailed guidance on the measures used to evaluate educators and on the design of the pay-for-performance bonuses. The 2010 grants required four components in performance-based compensation systems implemented in districts, as well as five core elements needed to support the initial and ongoing implementation of the compensation systems. Next, we summarize these four required components.

### Required Components of the Performance-Based Compensation Systems

1. **Measures of educator effectiveness.** Grantees were required to use a comprehensive, multiple-component measure of effectiveness for teachers and principals. The measures had to include student achievement growth and at least two observations of classroom or school practices. In addition, the evaluation had to give significant weight to student achievement growth—defined as the change in student achievement for an individual

[3] A full list of the 2010 TIF grantees, including a profile of their performance-based compensation systems, can be found at http://cecr.ed.gov.

student between two or more points in time. Only trained observers using objective, evidence-based rubrics could conduct the observations. Grantees had discretion to include additional measures.

2. **Pay-for-performance bonus.** Grantees were required to offer bonuses to educators based on how they performed on the effectiveness measures. The bonuses were designed to incentivize educators and to reward them for being effective in their classroom and schools. There were no additional requirements for earning the bonus beyond performing well on the effectiveness measure. To provide a strong incentive for the most effective educators, bonuses were to be differentiated and substantial enough to lead to change in the behavior of teachers and principals to improve student outcomes.
3. **Additional pay opportunities.** The performance-based compensation systems had to include pay opportunities for educators to take on additional roles or responsibilities. These roles might include becoming a master or mentor teacher who directly counsels other teachers or develops or leads professional development sessions for teachers. Limiting these additional pay opportunities to educators identified as effective could also provide an incentive for educators to improve their effectiveness. However, those educators would need to agree to take on leadership roles and perhaps work additional hours.
4. **Professional development.** TIF grantees were required to support teachers and principals in their performance improvement efforts. Support included providing information about measures on which educators would be evaluated and more targeted professional development based on an educator's actual performance on the effectiveness measures. Specifically, districts were required to provide educators with feedback and professional development on how to alter their pedagogy or practices to improve along the measures.

These four components of a performance-based compensation system were required of all grantees. In addition, ED encouraged the use of other components that would provide additional pay by awarding points to applicants that included these features in their performance-based compensation systems. For example, districts could offer additional pay to effective educators who agreed to work in hard-to-staff subjects, such as secondary math and science in high-need schools.

### Core Elements Designed to Support Implementation of the Performance-Based Compensation System

TIF grantees also were required to have the proper supports to implement and maintain the performance-based compensation system. The five core elements were (1) the involvement and support of teachers, principals, unions (if applicable), and other personnel needed to carry out the TIF grant; (2) a rigorous, transparent, and fair evaluation system for teachers and principals; (3) a plan to effectively communicate the components of the grantee's performance-based compensation system; (4) a plan for ensuring educators understood the measures of educator effectiveness; and (5) a data management system that could link student achievement data to educator payroll and human service systems (see Max et al. 2014 for more details on the core elements).

The required components of the performance-based compensation system are comprehensive and designed to work together, so grantees had to have the core elements in place before implementing their compensation systems. Grantees that did not have all the core elements in place when they were

awarded their grants in 2010 were required to spend the 2010–2011 school year planning and developing the support for implementation, and most grantees used the 2010–2011 school year as a planning year (Max et al. 2014). All grantees were required to begin implementation of their performance-based compensation systems by the 2011–2012 school year.

**Areas of Discretion in Performance-Based Compensation System Designs**

Although the TIF grant required grantees to include specific components in the performance-based compensation system, it gave them substantial discretion in designing and implementing these components. For example, grantees could assess a teacher's measured effectiveness based on the achievement growth of that teacher's students, all students in the same grade, the entire school, or some combination of these measures. Grantees could measure student achievement growth using a value-added model or by calculating the change in students' achievement on a standardized test from one year to the next. They could use models developed by the district, a vendor, or the state. Grantees could decide which rubrics they wanted to use to observe teachers and principals, the number of observations in a year (as long as there were at least two), and which staff members to train as observers. The criteria for earning a bonus based on the effectiveness measures also could vary (for example, criteria might require scoring above a predetermined threshold or in the top percentage on individual measures or a combination of measures). Grantees could choose bonus amounts based on educator performance. Finally, grantees could choose whether to offer retention and recruitment incentives (such as stipends) to educators to teach in high-need schools or to teach hard-to-staff subjects in those schools.

**Additional Requirements for Evaluation Grantees**

The 2010 TIF grant notice differed from the other rounds of the TIF grants in that it included a main competition and an evaluation competition (Max et al. 2014). By holding two separate competitions, ED created a sample of grantees that, by virtue of having applied for an evaluation grant, had indicated their interest and willingness to participate in a more in-depth evaluation of their TIF grants.

Evaluation grantees had to meet three additional grant requirements. First, they had to agree to participate in a random assignment evaluation of pay-for-performance bonuses. Schools within a district were randomly assigned to implement either all four required components of the performance-based compensation system program, including pay-for-performance bonuses (the treatment group), or all components *except* pay-for-performance bonuses (the control group). Second, evaluation grantees were required to include at least eight elementary or middle schools in the evaluation. Third, they were obligated to cooperate with all data collection activities for the evaluation.

Applicants for the evaluation grants were also given more specific guidance about the structure of their pay-for-performance bonus. They received examples of pay-for-performance bonuses that were *substantial* (with an average payout worth 5 percent of the average educator salary), *differentiated* (with at least some educators expecting to receive a payout worth three times the average payout), and *challenging* to earn (with only those performing significantly better than the average receiving bonuses). Although applicants had discretion over the proposed structure of the pay-for-performance bonus, these examples provided additional guidance to evaluation applicants and may have influenced how they designed their performance-based compensation systems.

In return for meeting the additional grant requirements, evaluation grantees received an extra $125,000 per school that participated in the evaluation. The money could be used to support the implementation of TIF—for example, to cover the cost of academic coaches or release time for professional development activities—as well as costs associated with the evaluation, such as data collection activities. The use of the funds also had to be consistent with the evaluation. For example, they could not be used to offer pay-for-performance in control schools.

ED monitored all grantees to ensure implementation was consistent with grant requirements. Although ED ensured all grantees received technical assistance, it used two providers—one for the non-evaluation grantees and one for the evaluation grantees. Resources for the evaluation grantee technical assistance team helped ensure that the evaluation grantees received intensive and targeted assistance. The evaluation grantee technical assistance team encouraged and supported evaluation grantees to incorporate criteria for their pay-for-performance bonuses consistent with their specific grant and in keeping with the examples provided in the grant notice. The goal of the technical assistance provided to all grantees was to ensure strong implementation that could bring about change in educational practices to improve student achievement, as specified in the logic model described below.

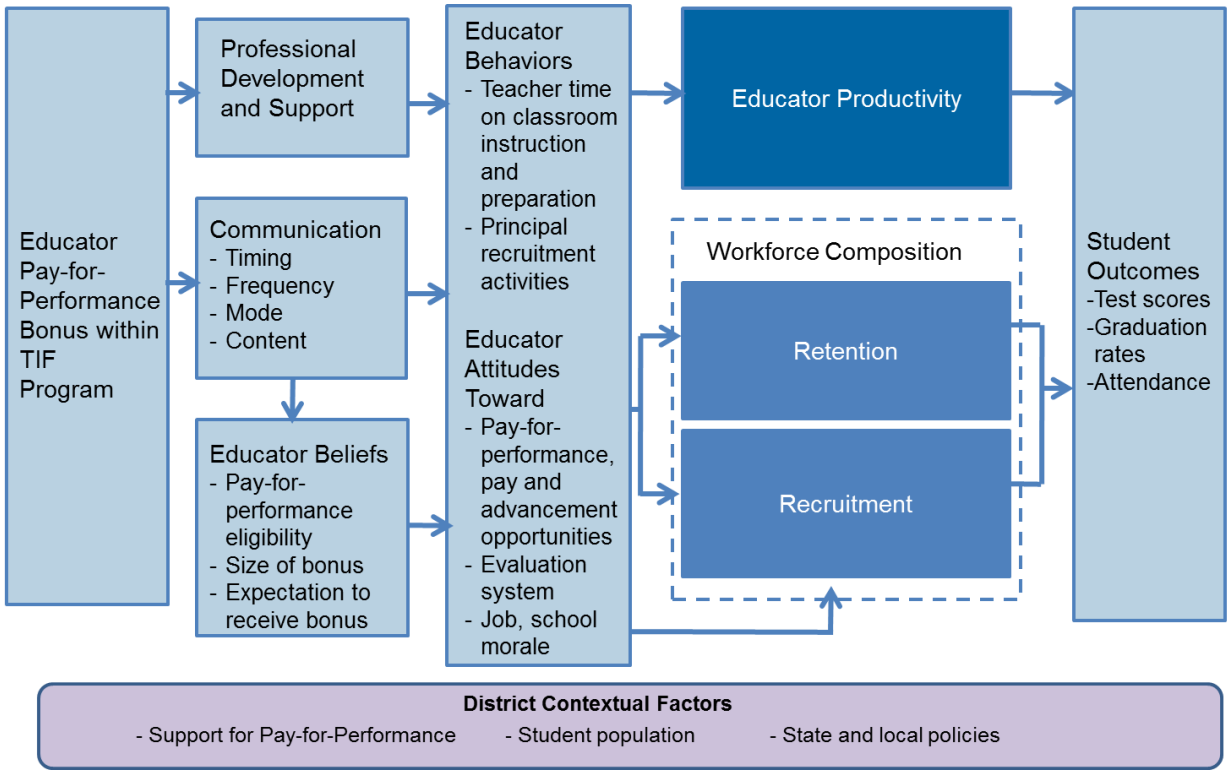
## Logic Model: How Pay-for-Performance Could Influence Student Outcomes

The requirements of the TIF grant, as well as the design of the evaluation of pay-for-performance bonuses, were informed by a theory of change of how pay-for-performance, within a comprehensive TIF performance-based compensation program, might lead to improved student outcomes. We developed a logic model to show the pathways by which the pay-for-performance component of TIF could influence student outcomes (Figure I.1). These pathways show the type of information needed to determine whether pay-for-performance is having a positive, negative, or neutral effect and thus informed the data collected as part of the evaluation.

As the starting point for the theory of change, districts adopt a TIF program that includes pay-for-performance bonuses for rewarding educators based on their measured effectiveness. The ability to earn a pay-for-performance bonus, as well as the fact that the criteria to earn a bonus depend on student achievement gains, could affect teachers' attitudes toward their school choice, alter their teaching practices, and increase their productivity. For example, pay-for-performance bonuses may serve as incentives for effective teachers to remain in a school that provides bonuses and may attract other effective teachers to the school. In addition, pay-for-performance bonuses based on schoolwide student achievement gains may encourage teacher collaboration, which may increase educator productivity. Educators rewarded for student achievement gains on standardized tests may allocate more time to instructional practices intended to improve test scores.

However, whether and how pay-for-performance bonuses actually lead to changes in educator productivity and the composition of the teaching workforce depends on many factors. For example, educators must be aware they are eligible to earn a bonus. Simply adopting a well-designed pay-for-performance program will not change teaching practices if educators do not know they are eligible. In addition, educators may be incentivized by pay-for-performance bonuses only if they understand how they are being evaluated and how they can change their teaching practices to improve their performance. They also must believe they are being evaluated consistently and fairly and that the bonuses are attainable and large enough to warrant changing their behavior. The critical role communication and professional development play in the logic model highlights the emphasis on these activities required by the grant.

**Figure I.1. Logic Model**




Educators' understanding of their TIF program will depend on districts' communication activities, timing of communication, and educators' receiving the information. Educators' awareness and understanding of the program can depend on the frequency, content, and types of district communication. Yet even a well-communicated program may be misunderstood if the program is complicated or if educators do not attend informational meetings or read the materials offered. Furthermore, educators must be made aware of the program when there is still sufficient time to affect their school choice (for example, request a school transfer) or to alter their teaching practices.

The ability of pay-for-performance bonuses to affect educator behaviors and attitudes also depends on the district context, such as educators' support for performance bonuses and the presence of other policies. If few educators in the school support pay-for-performance initiatives, adopting such a program may diminish school morale and job satisfaction, thereby decreasing productivity or inducing effective educators to leave the school.[5] District hiring policies, such as hiring freezes, may restrict mobility and negate potential benefits. Other existing policies, such as the requirements for teacher tenure, may already provide strong incentives for educators to improve student outcomes, diminishing the potential impact of performance bonuses. Finally, for schools at risk of closing because they have been designated as needing improvement, the introduction of a pay-for-performance program may not provide additional incentive for change.

[5] Many studies from the behavioral economics and psychology literature have examined how incentives and the design of incentive programs can affect behaviors. For example, some researchers have found that incentives may be ineffective or harmful if they decrease intrinsic motivation or are too weak, or if people believe they cannot meet the criteria to receive them. Others, however, have found that properly designed incentives can have a positive effect on productivity. See Kamenica (2012) for a review.

Even a well-designed and well-implemented comprehensive compensation reform program may take more than a year before it can have an impact on student achievement. For example, educators may not initially understand the incentives they are eligible to receive, know how to effectively change their teaching practices based on feedback provided through the district evaluation system, or be willing to change their behavior until they experience performance bonus payouts. Districts may need time to (1) design or revise performance measures so they can provide useful and accurate information to educators, (2) understand how to provide professional development that can help educators improve on the performance measures, and (3) effectively explain to educators how they are being evaluated and how bonuses are determined. It also may take time for the policy to cause changes in the overall quality of the educator workforce through the retention and recruitment of high quality teachers and principals. Because these learning and feedback processes may take multiple school years, it could take several years for impacts on student outcomes to be realized.

## Research Questions

The purpose of this multiyear study is to describe the program characteristics and implementation experiences of 2010 TIF grantees and estimate the impact of pay-for-performance bonuses within a well-implemented performance-based compensation system. Because educators' understanding of and response to this policy can change over time, the study plans to follow the grantees for the full duration of the grants.

The study will address four research questions:

1. What are the characteristics of all TIF districts and their performance-based compensation systems? What implementation experiences and challenges did TIF districts encounter?
2. How do teachers and principals in schools that did or did not offer pay-for-performance bonuses compare on key dimensions, including their understanding of TIF program features, exposure to TIF activities, allocation of time, and attitudes toward teaching and the TIF program?
3. How do pay-for-performance bonuses affect educator effectiveness and the retention and recruitment of high-performing educators?
4. What is the impact of pay-for-performance bonuses on students' achievement on state assessments in math and reading?

The first report from this study (Max et al. 2014) described implementation of TIF for all 2010 grantees and, for a subset of 10 evaluation districts, provided detailed findings on implementation and the effect of pay-for-performance bonuses on educators' reported satisfaction, attitudes, and behaviors. This report found that fewer than half of all 2010 TIF districts reported implementing all four required components of their TIF program. For the 10 evaluation districts, the report indicated that (1) many educators misunderstood the measures used to evaluate their performance, their eligibility for a pay-for-performance bonus, and the potential amount of the performance bonus they could earn; (2) most educators were satisfied with their professional opportunities, school environment, and the TIF program; and (3) educators in schools that offered pay-for-performance bonuses tended to be less satisfied than those in schools that did not offer performance bonuses.

This second report focuses on implementation and the effect of pay-for-performance in the 10 evaluation districts after one and two years of program implementation. It captures the views and

attitudes of educators based on their early experiences with pay-for-performance bonuses—once before and once after they were aware of how they performed on the measures of effectiveness for the 2011–2012 school year and what bonus, if any, they received for that performance. The report also presents early impacts of pay-for-performance on educator effectiveness and student achievement. These analyses are based on information obtained from educator and district surveys, interviews with TIF district administrators, and student and educator administrative data provided by the evaluation districts. Although the report focuses on the 10 evaluation districts, it also includes information on implementation of TIF for all 2010 grantees.

## Road Map for the Remainder of the Report

In the rest of this report, we describe in detail the study's design and findings. In Chapter II, we describe the study sample, design of the experimental evaluation, data used for this report, and analytic approaches. In Chapter III, we describe the programs of all 2010 TIF districts and challenges the districts encountered in implementing TIF. In Chapter IV, we provide more detailed information on implementation experiences in TIF evaluation districts, and, in Chapter V, we examine the impact of eligibility for pay-for-performance bonuses on teachers' and principals' attitudes and behaviors. Finally, in Chapter VI, we present findings on the impact of pay-for-performance on educator effectiveness and student achievement.

THIS PAGE IS INTENTIONALLY BLANK

# II. STUDY SAMPLE, DESIGN, DATA, AND METHODS

In this chapter, we describe the study sample, design, and data used for this report. We also present an overview of the analytic approaches.

## Study Sample

This study is based on school districts and schools that were part of the Teacher Incentive Fund (TIF) grants awarded in 2010 by the U.S. Department of Education (ED). That year, ED awarded 62 TIF grants that included 183 districts. As we explained in Chapter I, the 2010 grants were awarded under two separate competitions: (1) a main competition; and (2) an evaluation competition, for which grantees agreed to participate in a study that involved random assignment of schools to a treatment group or a control group. Most of this report focuses on the TIF districts that were part of the evaluation competition, which we refer to as "evaluation districts."[6] We refer to the remaining TIF districts as "non-evaluation districts."

Most, but not all, districts in the 2010 grants participated in TIF in subsequent years. A total of 171 districts implemented TIF—that is, had a performance-based compensation system supported by TIF funds—in 2011–2012, and 164 districts implemented TIF in 2012–2013 (Table II.1).[7] Among the districts that implemented TIF in 2012–2013, 13 were evaluation districts: a new district and the 12 districts that had also implemented TIF the previous year.

**Table II.1. Number of Districts Implementing TIF, by Year**

| | Implemented TIF in 2011–2012 | Implemented TIF in 2012–2013 | Responded to 2013 District Survey |
|---|---|---|---|
| Non-evaluation districts | 159 | 151 | 142 |
| Evaluation districts | 12 | 13 | 13 |
| **Total** | **171** | **164** | **155** |

Source: U.S. Department of Education and TIF grantee reports.

Note: A district is regarded as implementing TIF if it had at least some components of a performance-based compensation system supported by TIF funds. Districts that had a TIF program in both 2011–2012 and 2012–2013 are included in the counts for both years.

Districts were awarded, or included in, a TIF grant through a competitive process, and the grants were designed to serve high-need schools. Therefore, TIF districts were not representative of all U.S. districts. An earlier report from this study (Max et al. 2014) showed that, compared to the average U.S. district, TIF districts were larger, were more likely to be urban and located in the South, and had a higher proportion of students who were racial/ethnic minorities and eligible for free or reduced-price lunch.

[6] For this study, one set of charter schools that were part of the same TIF evaluation grant, were in the same state, and belonged to a common charter school association was considered to be a single evaluation district.

[7] Between 2011–2012 and 2012–2013, eight non-evaluation districts withdrew from their grants, and one evaluation grantee added a district to its TIF grant.

This report provides an overview of TIF implementation in all TIF districts in 2012–2013 and an in-depth analysis of implementation and the impacts of pay-for-performance on educator and student outcomes in the evaluation districts. Next, we describe the final sample of districts included in these analyses.

### All TIF Districts in the Final Analysis Sample

In Chapter III of this report, we examine TIF implementation in all TIF districts (evaluation and non-evaluation) in the 2012–2013 school year—the second year of implementation for nearly all those districts. We describe the districts' reported compliance with implementing the four required components of TIF and the challenges they encountered in implementing TIF. As discussed later, this analysis relied on districts' responses to a survey we administered in 2013. Therefore, the final sample for this analysis consisted of 155 TIF districts—13 evaluation and 142 non-evaluation districts—that participated in TIF in 2012–2013 and responded to the district survey (Table II.1).

### Evaluation Districts in the Final Analysis Sample

The rest of this report focuses on the evaluation districts, from which we collected more detailed information. This information—obtained from surveys, interviews, technical assistance documents, and administrative data—allowed us to describe the performance bonuses and performance ratings that educators actually earned, document districts' strategies for communicating key program features, analyze educators' understanding of and attitudes toward TIF, and estimate the impact of pay-for-performance on educator and student outcomes.

ED used the same criteria to award evaluation and non-evaluation TIF grants, but evaluation districts may differ from other TIF districts in important ways related to the evaluation requirements. The requirement to provide at least eight elementary or middle schools for the evaluation may have resulted in larger districts being part of the in-depth evaluation. In addition, the requirement for random assignment of pay-for-performance bonuses may have drawn in districts that were confident they could obtain educator buy-in to randomly assign this required program component.

Evaluation and non-evaluation districts differed on several demographic and socioeconomic characteristics (Table II.2). Although we found few statistically significant differences, the relatively small sample size of 13 evaluation districts implied that only large differences would have been statistically significant. Therefore, we note differences that were larger than 10 percentage points or 10,000 students. Evaluation districts were larger, on average, than non-evaluation districts. Evaluation districts were also more likely than non-evaluation districts to be in urban areas (69 versus 33 percent) and in the West (46 versus 17 percent), and less likely to be in rural areas (8 versus 33 percent), in the South (23 versus 47 percent), and in states with collective bargaining agreements (54 versus 71 percent). Evaluation and non-evaluation districts had similar proportions of students who were black or Hispanic or that received free or reduced-price lunch.

**Table II.2. Comparison of TIF Evaluation Districts and Non-Evaluation Districts (Percentages Unless Otherwise Noted)**

| | Evaluation Districts | Non-Evaluation Districts |
|---|---|---|
| Student Racial/Ethnic Distribution | | |
| White, non-Hispanic | 39 | 49 |
| Black, non-Hispanic | 33 | 25 |
| Hispanic | 21 | 19 |
| Student Socioeconomic Status | | |
| Eligible for free/reduced-price lunch | 63 | 63 |
| Title 1 eligible schools (schoolwide) | 66 | 76 |
| Enrollment (Average) | | |
| Number of students | 32,450 | 19,496 |
| District Location | | |
| Urban | 69 | 33* |
| Suburban | 8 | 14 |
| Town | 15 | 20 |
| Rural | 8 | 33* |
| Geographic Region | | |
| Northeast | 15 | 8 |
| Midwest | 15 | 28 |
| South | 23 | 47 |
| West | 46 | 17* |
| Collective Bargaining[a] | | |
| In state with collective bargaining | 54 | 71 |
| **Number of States** | **8** | **24** |
| **Number of Districts** | **12-13** | **144-156** |

Source: Common Core of Data for 2011–2012 school year.

Notes: The table is based on all 169 districts that had a TIF program in the 2012–2013 school year. Ten non-evaluation districts were not included in the 2010–2011 district-level data from the Common Core of Data. Common Core of Data school-level data are used to calculate socioeconomic indicators. Common Core of Data district-level data are used to calculate all other demographic characteristics.

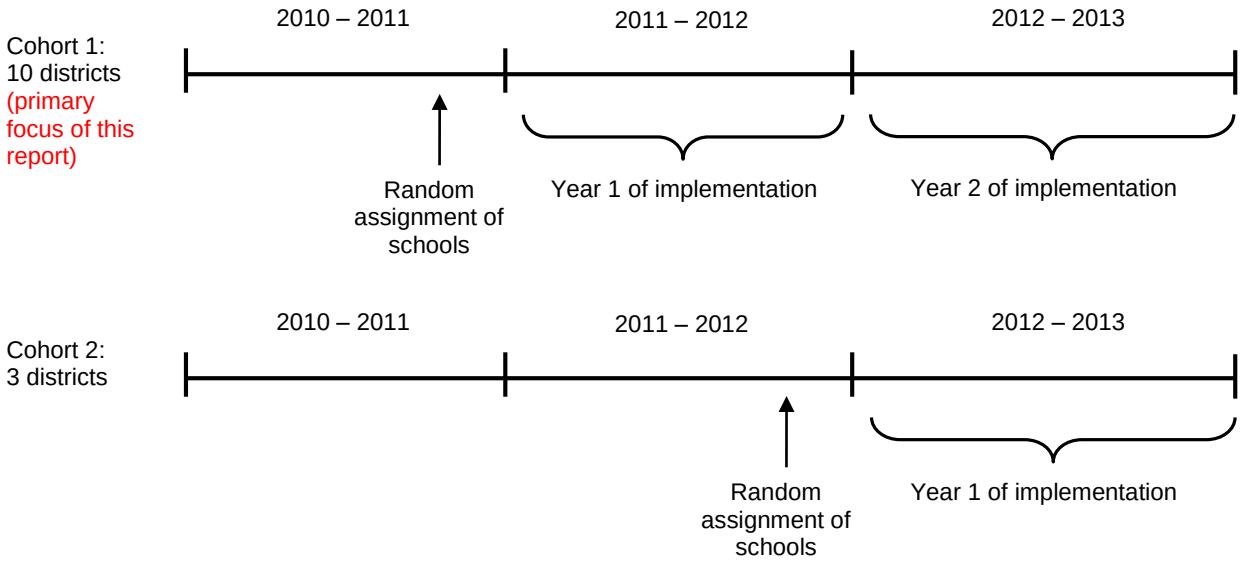
[a]Collective bargaining is a state-level indicator from the National Right to Work Legal Defense Foundation (http://www.nrtw.org/rtws.htm).

*Difference between evaluation and non-evaluation districts is statistically significant at the .05 level, two-tailed test.

We classified evaluation districts into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group and a control group (Figure II.1). Cohort 1 consisted of 10 districts for which we randomly assigned schools in spring and summer 2011. From these districts, we obtained data on two years of TIF implementation: 2011–2012 (Year 1) and 2012–2013 (Year 2). Cohort 2 consists of three districts for which we randomly assigned schools in spring and summer 2012 and obtained data on one full year of TIF implementation, 2012–2013, representing Year 1 of this cohort's implementation of TIF.[8]

[8] Two Cohort 2 districts began putting some components of their TIF programs into place in 2011–2012, and Table II.1 includes these two districts in the counts of districts that implemented TIF in 2011–2012. However, because these districts were not ready for random assignment of schools until spring and summer 2012, we classified them as Cohort 2 districts and, for this report, specified 2012–2013 as Year 1 of the districts' implementation of TIF.

**Figure II.1. Two Cohorts of Evaluation TIF Districts**




The structure of the grants varied among the 10 Cohort 1 districts. Four of these districts received TIF grants directly from the U.S. Department of Education. The remaining six Cohort 1 districts were part of multidistrict grants that were administered by another grantee organization—such as a state education agency, university, association of charter schools, or nonprofit organization. In total, the 10 Cohort 1 districts represented eight distinct grantees.

**This report primarily focuses on the 10 Cohort 1 evaluation districts—those for which data were available on two years of TIF implementation.** We refer to the first and second years of implementation for Cohort 1, 2011–2012 and 2012– 2013, as Years 1 and 2. As explained in Chapter I, because TIF is a comprehensive program for reforming educator compensation and improving educator effectiveness, it may take time for educators to fully understand the incentives available, the measures on which they are evaluated, and the improvements they need to make to earn bonuses. An earlier report from this study (Max et al. 2014) presented findings on educators' understanding, attitudes, and behaviors from Year 1 of TIF implementation—before performance ratings were determined and bonuses were distributed. Educators' perceptions and practices may have changed after they experienced the results of the performance evaluations and bonuses and determined how to respond to this new information. Focusing on Cohort 1 districts allowed us to examine such changes between Years 1 and 2 and assess whether impacts on educator and student outcomes also evolved between years, while ensuring the same schools were included in the analysis for both years. Unless otherwise noted, all findings in Chapters IV through VI are based on these 10 Cohort 1 districts.[9]

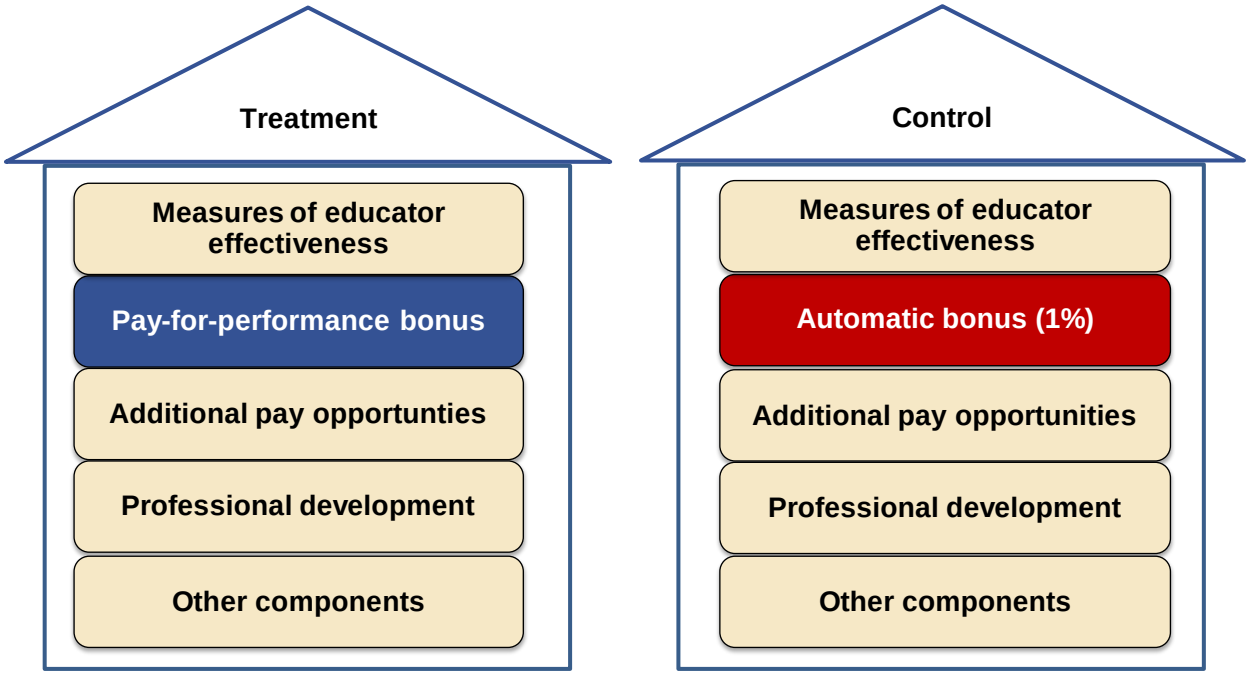
## Experimental Design to Estimate the Impact of Pay-for-Performance

To ensure that the study's findings on the impacts of pay-for-performance could be attributed solely to the offer of pay-for-performance and not to other characteristics of districts, schools, or educators, we randomly assigned elementary and middle schools within each district to treatment and

[9] For key implementation features and outcomes, the appendices of this report provide findings from Year 1 of implementation for Cohorts 1 and 2 together—that is, findings from 2011–2012 for Cohort 1 and from 2012–2013 for Cohort 2.

control groups. In Figure II.2, we illustrate the experimental design and highlight that treatment and control schools were expected to implement the same features of the district's performance-based compensation system, except for the pay-for-performance component. Educators (teachers and principals) at treatment schools were eligible to earn a pay-for-performance bonus; educators at control schools received an automatic bonus worth approximately 1 percent of their salary each year. The 1 percent bonus ensured that all educators in evaluation schools received some benefit from participating in the study: either the opportunity to earn a pay-for-performance bonus or the automatic bonus. Therefore, the impact of pay-for-performance estimated in this study potentially reflects two key differences between treatment and control schools: (1) bonuses in treatment schools were differentiated based on performance; and (2) bonuses in treatment schools were larger, on average, than in control schools.

**Figure II.2. Random Assignment Design**




Evaluation districts chose which schools would be included in the evaluation. Because a primary objective of the study was to measure the impact of pay-for-performance on student achievement on state assessments in high-need schools, every participating school needed to have (1) at least half of its students receiving free or reduced-price lunch, and (2) at least one grade level tested by state assessments (3rd to 8th grade).

Before random assignment, schools were paired based on having similar characteristics measured before the district's implementation of TIF—primarily student achievement, grade span, and school size. District staff either approved the pairs we constructed or directly specified the pairs based on their knowledge of the participating schools. One school from each pair was randomly assigned to the treatment group, and the other school in the pair was assigned to the control group. We describe random assignment procedures in more detail in Appendix A.

We randomly assigned 183 elementary and middle schools to either the treatment or control group—138 schools assigned as part of Cohort 1 and 45 additional schools as part of Cohort 2 (Table II.3). Of the 138 Cohort 1 schools, our primary analysis sample consisted of 132 schools that

implemented the TIF program for two years.[10] This sample excluded schools that closed or dropped out of the study along with the schools with which they were paired—a total of six schools (4 percent of all Cohort 1 schools). Appendix A, Table A.1 describes this school attrition in more detail.

**Table II.3. Number of Schools in the Evaluation, by Cohort and Treatment Status**

| Cohort (# districts) | Timing of Random Assignment | Number of Treatment Schools | Number of Control Schools | Total Number of Schools |
|---|---|---|---|---|
| Cohort 1 (10 districts) | Spring/summer 2011 | 69 | 69 | 138 |
| Cohort 2 (3 districts)[a] | Spring/summer 2012 | 23 | 22 | 45 |
| **Number of Schools** | | **92** | **91** | **183** |
| **Final Analysis Sample (Schools in Cohort 1 that implemented TIF for 2 years)** | | **66** | **66** | **132** |

[a]Counts of schools that were randomly assigned in spring/summer 2012 include a small number of schools (fewer than 3) from Cohort 1 districts to replace schools that closed.

## Baseline Characteristics of Treatment and Control Schools

The key advantage of this study's random assignment design is that, at the beginning of the study, the treatment and control groups were expected to include students and educators with similar characteristics. Because the two groups were expected to differ only in the opportunity for educators to receive pay-for-performance bonuses, differences in outcomes between the groups could be rigorously attributed to the impact of pay-for-performance.

At the beginning of the study, we found that treatment and control schools in the final analysis sample were similar on most of the measured characteristics of their students and educators. In the pre-implementation year—the year of random assignment before the first year of TIF implementation—the overall difference in student characteristics between treatment and control schools was not statistically significant ($p$=0.096; Table II.4). On a few specific student characteristics, treatment and control schools differed slightly. Students in treatment schools had slightly lower achievement in math than students in control schools (the difference did not exceed 0.05 standard deviations). In addition, the percentage of students who were white was lower in treatment schools than in control schools by 3 percentage points. Treatment and control schools had similar student achievement in reading before the implementation of TIF and similar fractions of students who received free or reduced-price lunch, had an Individualized Education Program, were overage for their grade, or were English language learners. As discussed later in this chapter, all analyses of the impacts of pay-for-performance on educator and student outcomes were adjusted to account for the slight preexisting differences in student achievement and racial/ethnic composition between treatment and

[10] Analyses that used administrative data were based on all 132 schools. Analyses that used educator survey data were based on 131 schools in 2011–2012 and 132 schools in 2012–2013. When we administered the spring 2012 educator surveys, we did not know that one school was a multicampus school with different administrative structures, and therefore only one of the campuses was surveyed.

control schools. Treatment and control schools had similar educator characteristics in Year 1, the first year of educator data available for all districts (Table II.5).[11,12]

**Table II.4. Characteristics of Students Enrolled in Treatment and Control Schools in the Pre-Implementation School Year (2010–2011) (Percentages Unless Otherwise Indicated)**

| | Treatment | Control | Difference |
|---|---|---|---|
| Achievement in the Pre-Implementation Year (*z*-score) | | | |
| Math | -0.47 | -0.43 | -0.04* |
| Reading | -0.41 | -0.40 | -0.02 |
| Race/Ethnicity | | | |
| White, non-Hispanic | 27 | 30 | -3* |
| African American, non-Hispanic | 44 | 42 | 2 |
| Hispanic | 23 | 22 | 1 |
| Other | 6 | 6 | -1 |
| Other Characteristics | | | |
| Female | 49 | 49 | -1 |
| Eligible for free/reduced-price lunch | 77 | 76 | 1 |
| Disabled or has an Individualized Education Program | 12 | 12 | 0 |
| Overage for grade | 13 | 13 | 0 |
| English language learner | 9 | 9 | 0 |
| Grade Span | | | |
| Grades 3-5 | 64 | 64 | 0 |
| Grades 6-8 | 36 | 36 | 0 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.096 |
| **Number of Students—Range[a]** | **12,640-22,163** | **12,523-22,065** | |
| **Number of Schools—Range[a]** | **42-66** | **42-66** | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

The study schools' baseline characteristics confirm that the schools were both high-need and low-performing. As Table II.4 shows, in both the treatment and control schools, at least three-fourths of the students received free or reduced-price lunch, and the students' math and reading achievement was lower than the average achievement in their states by at least four-tenths of a standard deviation.

[11] Appendix A, Tables A.2 and A.3 show the characteristics of all study schools in Cohorts 1 and 2 at the beginning of the study. We found that treatment and control schools in this sample were similar on most of the measured characteristics of their students and educators.

[12] Appendix A, Table A.4 shows educator characteristics within treatment and control schools in the pre-implementation year for 9 of 10 districts that provided educator data for that year. In these districts, treatment and control schools were similar on most of the characteristics of their educators, with a few exceptions: teachers in treatment schools were 3 percentage points more likely than those in control schools to be white and 3 percentage points less likely to be black.

**Table II.5. Characteristics of Educators in Treatment and Control Schools in Year 1 (Percentages Unless Otherwise Noted)**

| | Teachers | | | Principals | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 85 | 84 | 1 | 62 | 56 | 6 |
| Race/ethnicity | | | | | | |
| White, non-Hispanic | 74 | 72 | 2 | 59 | 56 | 3 |
| Black, non-Hispanic | 19 | 21 | -2 | 32 | 36 | -4 |
| Hispanic | 2 | 2 | 0 | 3 | 2 | 2 |
| Other | 5 | 4 | 0 | 6 | 7 | -1 |
| Age (average years) | 42 | 41 | 0 | 49 | 48 | 1 |
| Education | | | | | | |
| Master's degree or higher | 51 | 50 | 0 | 94 | 93 | 1 |
| Experience in K-12 education | | | | | | |
| Total experience (average years) | 12 | 11 | 0 | 16 | 15 | 2 |
| Less than 5 years | 25 | 25 | 0 | 18 | 14 | 4 |
| 5-15 years | 45 | 46 | -2 | 34 | 40 | -6 |
| More than 15 years | 30 | 28 | 2 | 48 | 46 | 2 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.745 | | | 0.516 |
| **Number of Educators—Range[a]** | **1,458-2,076** | **1,500-2,055** | | **40-59** | **45-65** | |
| **Number of Schools—Range[a]** | **49-66** | **49-66** | | **38-57** | **43-61** | |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

## Data Sources

The analyses in this report are based on data from eight sources. Table II.6 summarizes the data sources, along with response rates. Next, we describe each of these data sources in more detail.

### Data for 2010 TIF Districts

**Common Core of Data.** This publicly available database provided information on the characteristics of all TIF districts, including students' race and ethnicity, free or reduced-price lunch eligibility, average district enrollment, and geographic information. We used data from the 2010– 2011 school year to compare the characteristics of evaluation and non-evaluation districts.

**District survey.** The district survey asked TIF districts to provide information on the components of their TIF programs, program communication strategies, and general experiences and challenges in implementation. We addressed these surveys to the person identified as overseeing or directing each district's TIF program. Districts' responses allowed us to describe programs in all TIF districts and to determine their compliance with the four required components of the TIF grant.

**Table II.6. Data Sources for This Report**

| Data Source | Type of Information | Response Rates (Percentages) 2011–2012 | 2012–2013 |
|---|---|---|---|
| **Data Collected from Evaluation and Non-Evaluation Districts** | | | |
| 1. Common Core of Data | Composition of student characteristics in districts | NA | NA |
| 2. District survey | TIF program features, implementation experiences | 91 | 95 |
| **Data Collected from Evaluation Districts Only** | | | |
| 3. District interviews | Detailed information on TIF implementation and program features | 100 | 100 |
| 4. Principal survey | TIF program features, attitudes toward TIF program and job, hiring practices | 98 | 95 |
| 5. Teacher survey | TIF program features, attitudes toward TIF program and job, time use | 92 | 92 |
| 6. Technical assistance documents | Detailed information on implementation and program features | 100 | 100 |
| 7. Student administrative data | Students' standardized test scores and background characteristics (grades 3 through 8) | 100 | 100 |
| 8. Educator administrative data | Teachers' and principals' school assignments, background characteristics, performance ratings, and compensation from TIF | 100 | 100 |

Note: Survey response rates are shown for treatment and control groups combined. None of the response rates differed between the treatment and control groups by a statistically significant margin or by more than 3 percentage points.

NA is not applicable.

We administered the survey in both 2012 (in the middle of the 2011–2012 school year) and 2013 (near the end of the 2012–2013 school year) to all districts participating in TIF in those years. This report primarily used data from the 2013 survey to describe the programs in 2012–2013; in some cases, however, we used data from the 2012 survey to examine whether compliance with required components changed from the first year to the second year of TIF implementation. In 2013, 95 percent of TIF districts responded to the district survey (Appendix A, Table A.5). Districts that responded and did not respond to the survey did not differ by a statistically significant margin on most characteristics—including the districts' student racial composition, student socioeconomic status, and size—but respondents were more likely to be in urban and rural areas than suburban areas, and more likely to be in the West region (Appendix A, Table A.6).

### Data for TIF Evaluation Districts Only

**District interviews.** Interviews with TIF program administrators in evaluation districts provided more in-depth information than that collected from the survey. Through these interviews, we probed for more details on how bonuses were determined, how the program was communicated to educators, the timing of bonus awards, types of challenges encountered in implementation, and revisions to the program to overcome those challenges. Information from the interviews allowed us to develop a

comprehensive description of implementation in evaluation districts and, when appropriate, to fill in missing information or supplement survey responses. This report used data from the first and second years of interviews, which we conducted with all 13 evaluation districts participating in TIF in summer 2013.

**Principal and teacher surveys.** We administered surveys to principals and teachers in the evaluation districts to learn about their understanding of and experiences with TIF program components, job satisfaction, attitudes toward TIF, and job-specific practices (such as principals' approaches to hiring teachers and teachers' allocation of time). We used educator survey responses for three main purposes: (1) to describe educators' understanding of their TIF program; (2) to compare the experiences, attitudes, and classroom and school practices of educators in treatment and control schools; and (3) to examine how educators' understanding and attitudes may have changed.

In spring 2012 and spring 2013, we administered surveys to all principals and a sample of teachers within treatment and control schools that were participating in TIF in those years. Among full-time teachers, the teacher sample included all 4th-grade teachers; all 7th-grade math, English/language arts, and science teachers; and 77 percent of 1st-grade teachers in 2012 and 100 percent of 1st-grade teachers in 2013. These groups represent elementary and middle school grades and subjects both with and without annual accountability testing.[13]

Among the Cohort 1 districts, 98 percent of principals and 92 percent of teachers responded to the 2012 survey, and 95 percent of principals and 92 percent of teachers responded to the 2013 survey (Appendix A, Table A.7).[14] There were no statistically significant differences in response rates between treatment and control educators. Moreover, we found few differences between the characteristics of respondents and nonrespondents to the teacher survey (Appendix A, Table A.10).[15] Among both teachers and principals, we found few differences between the characteristics of respondents from treatment and control schools (Appendix A, Tables A.11 and A.12).

**Technical assistance documents.** The technical assistance team documented aspects of the evaluation districts' programs and implementation activities and experiences.[16] The team conducted needs assessments in fall 2010 and spring 2011 for each evaluation district or grantee. The assessments examined evaluation districts' program design and planned implementation, progress in implementing the five core elements required by ED, and use of communication materials during the planning year to inform educators about the program.

The evaluation team reviewed the documents for all evaluation districts. When appropriate, the team used this information to report more detail on the evaluation districts' TIF programs and implementation experiences.

[13] In 2013, we also surveyed teachers from the 2012 sample even if they left teaching, left the study schools, or switched teaching assignments. These teachers were not included in the final analysis sample. In Appendix A, we explain in detail how we determined the teacher sample.

[14] Appendix A, Table A.8 provides response rates for Cohort 2, and Table A.9 shows the distribution of grade and subject assignments for the Cohort 1 teachers who responded to the survey and were included in the final analysis sample.

[15] We do not report comparisons of respondents and nonrespondents to the principal survey due to the small number of nonrespondents.

[16] The technical assistance team consisted of Mathematica staff and consultants from Vanderbilt University.

**Student administrative data.** We collected evaluation districts' administrative records on students enrolled in treatment and control schools. The data included information on students' background characteristics and their scores on state assessments in math and reading, allowing us to examine the impact of pay-for-performance on student achievement. Within Cohort 1 districts—those that completed two years of TIF implementation—the data covered all students in study schools in 2010–2011 to 2012–2013, representing the period from the pre-implementation year to Year 2 of implementation. We obtained similar data from Cohort 2 districts for 2011–2012 and 2012–2013.

**Educator administrative data.** We collected evaluation districts' administrative records on teachers and principals, including information on their assignments to schools, background characteristics, performance ratings determined by their TIF programs, and compensation received from TIF. These data allowed us to describe thoroughly the performance ratings, bonuses, and additional pay that educators received from TIF and to examine the impact of pay-for-performance on educators' effectiveness and the retention and recruitment of effective educators.

Data on educators' school assignments covered a longer period than the other data used in this study. Within Cohort 1 districts, we collected the educator rosters of all study schools in the fall of each of the four school years from 2010–2011 to 2013–2014—the period from the fall of the pre-implementation year to the fall of Year 3 of implementation.[17] In our analysis of educator retention, school assignment data from the fall of Year 3 enabled us to determine whether educators returned to their school after the completion of Year 2. In each year after the pre-implementation year we also obtained the school assignments of educators who worked in a study school in the previous year but moved to a nonstudy school in the same district. We obtained similar data from Cohort 2 districts for 2011– 2012 to 2013–2014.

The other types of educator administrative data covered a period similar to that of the other data used in this report. The data included educators' background characteristics from 2010–2011 to 2012–2013 and their performance ratings and TIF-funded compensation from 2011–2012 and 2012–2013.

## Overview of Analytic Approach

In this section, we discuss the analytic approaches used in the rest of this report. Appendix B provides more technical details on the analytic methods.

### Implementation of TIF in All Districts (Chapter III)

To describe implementation in all 2010 TIF districts, presented in Chapter III, we drew primarily from district survey responses. For each measure of program implementation included on the district survey, our basic analytic approach was to calculate means or percentages, as appropriate. We gave each district equal weight so that findings reflected the experiences of the average district that implemented a TIF program.

### Implementation of TIF in Evaluation Districts (Chapter IV)

In Chapter IV, we describe the implementation of TIF in the 10 Cohort 1 districts that completed two years of program implementation. In addition to the district survey, we used information collected only from the evaluation districts: district interviews, technical assistance documents, administrative

[17] Four Cohort 1 schools in Year 1 and three in Year 2 did not have full-time principals (Appendix A, Table A.13). These schools were not included in the analysis of impacts on principals' outcomes measured from administrative data.

data on educators' performance ratings and compensation from TIF, and teacher and principal surveys.

To describe districts' program designs and implementation experiences, we used districts' responses to surveys and interviews to calculate means (or percentages, as appropriate), weighting each district equally. To describe actual bonus amounts and performance ratings, we used administrative data to calculate summary statistics (means, maximum levels, or percentages of educators receiving particular amounts or ratings) separately for each district and then took the equal-weighted average across all districts.

To describe educators' understanding of and experiences with TIF program components, we summarized educators' survey data separately by treatment status and year, giving each school equal weight. We compared the responses of treatment and control educators to determine whether they differed in their perceived eligibility for the component—pay-for-performance bonuses—that was supposed to differ between the two groups and whether they reported similar exposure to other components that were not supposed to differ. To ensure that any reported differences between the two groups were due solely to their differing eligibility for pay-for-performance rather than preexisting differences in the characteristics of their schools, we used a regression to adjust educators' reports for slight differences in baseline school characteristics in the same manner as done in our impact analyses, described below.

Educators' understanding of program components may change as they gain more exposure to those components. We examined how educators' understanding changed from Year 1 of TIF implementation (when no educators had yet received bonuses) to Year 2 (when educators had already seen one round of bonuses). Separately for treatment and control schools, we compared average reports in Year 1 and Year 2 and conducted hypothesis tests to determine whether differences between years were statistically significant.

**Impacts of Pay-for-Performance on Educator and Student Outcomes (Chapters V and VI)**

We estimated the impacts of pay-for-performance on several outcomes within the Cohort 1 evaluation districts. In Chapter V, we present impacts on educators' attitudes (such as job satisfaction) and self-reported behaviors (such as teachers' allocation of time and principals' hiring practices). In the theory of change in Chapter I, these attitudes and behaviors are intermediate factors that shape the key outcomes of interest: educator effectiveness (including the retention and recruitment of effective educators) and student achievement. In Chapter VI, we report the impacts of pay-for-performance on those key outcomes.

Because the study used random assignment, any differences in educators' or students' outcomes between the treatment and control group could be attributed to pay-for-performance and not some other characteristic of the districts or schools. We estimated these differences using a linear regression that accounted for the random assignment design—in particular, the assignment of schools rather than individuals to the treatment and control groups, as well as the pairing of schools before random assignment. As shown earlier in this chapter, treatment and control schools differed slightly in average student achievement and students' racial/ethnic composition before TIF implementation. Therefore, all regressions in the impact analyses accounted for the baseline differences by controlling for school averages of those student characteristics from the pre-implementation year. In some analyses, we also controlled for the individual characteristics of students or educators in the analysis samples to enhance

precision (see Appendix B for a full description of these characteristics).[18] We estimated regressions separately by year and used weights for educators' or students' data to give each school equal weight, so that the estimates reflected the impact of pay-for-performance on an average study school after one and two years of TIF implementation.[19]

Next, we discuss how we measured each type of outcome and determined the individuals whose outcomes were included in the impact analyses.

**Educators' attitudes and behaviors.** We measured educators' attitudes and self-reported behaviors directly from the survey responses of principals and teachers working in the study schools at the time of the survey administration. Analyses of teacher-reported outcomes were based on teachers who reported teaching 1st grade; 4th grade; or 7th-grade math, English/language arts, or science.

**Educator effectiveness.** We examined the impact of pay-for-performance on several measures districts used to evaluate educator effectiveness: (1) ratings based on the achievement growth of all students in a school (school achievement growth), which were used to evaluate both teachers and principals; (2) teachers' classroom observation ratings; (3) ratings based on the achievement growth of students in teachers' own classrooms (classroom achievement growth); and (4) observation ratings for principals. In each year, we examined impacts on the performance ratings of all full-time teachers and principals working in the study schools.[20]

**Retention and recruitment of effective educators.** On each performance measure, pay-for-performance could lead to higher average ratings by either enabling schools to retain and recruit more effective educators or motivating educators to improve their performance. Therefore, we also examined specifically whether pay-for-performance triggered staffing changes that increased the retention and recruitment of effective educators.

To assess whether pay-for-performance enabled schools to keep more effective educators, we examined differences between treatment and control schools in the effectiveness of *both* educators who stayed and those who left. If pay-for-performance caused more higher-performing educators to stay and more lower-performing educators to leave, then effectiveness should be higher among

[18] In this report, we present the average outcomes for the treatment group as regression-adjusted means. That is, we present the raw (unadjusted) average outcomes for the control group, and we compute the regression-adjusted treatment group mean as the sum of the control group mean and the estimated impact.

[19] The estimation of impacts in this report, described in Appendix B, differed slightly from methods used for the earlier report from this study (Max et al. 2014). For example, to estimate impacts for this report, we were able to use student administrative data to take into account differences between students' characteristics in treatment and control schools in the pre-implementation year. This led to slightly different estimates for Year 1 than those in the first TIF report (Max et al. 2014).

[20] Appendix B includes an explanation of how educator performance ratings were standardized. Appendix A, Tables A.14 and A.15 show the percentages of educators who received performance ratings; Tables A.16 through A.18 show the characteristics of educators who did and did not receive performance ratings; and Tables A.19 through A.21 compare the characteristics of educators in treatment and control schools who received performance ratings. We found few differences between the characteristics of educators with and without observation ratings, but teachers who received classroom achievement growth ratings in Year 2 were younger and less experienced than those who did not. Although there were few differences between the characteristics of treatment and control educators who received observation ratings, we found more differences between the characteristics of treatment and control teachers who received classroom achievement growth ratings; in particular, treatment teachers were older and more experienced.

educators who stayed in treatment schools than those who stayed in control schools, and lower among educators who left treatment schools than those who left control schools. We used performance ratings in Year 1 to measure the effectiveness of those who subsequently chose to stay in or leave their school.[21] Teachers and principals who worked in study schools were considered retained if they continued teaching in or leading the same school in Year 3 based on districts' administrative records.

To examine whether pay-for-performance enabled schools to recruit more effective teachers and principals to fill vacancies, we compared the Year 2 performance ratings of treatment and control educators who were new to their school in that year.

**Student achievement.** We measured student achievement using students' scores on state assessments in math and reading.[22] Because student achievement was measured on different scales in different states and grades, we standardized all scores into $z$-scores by subtracting the statewide grade-specific mean and dividing by the statewide grade-specific standard deviation. The analysis used all students in grades 3 through 8 who were tested in a study school in a given year. The tested students included those who had been enrolled in the same school at the time of random assignment and stayed in that school, as well as students who moved into a study school after random assignment.[23] Therefore, this analysis measured the impact of pay-for-performance on schools' average student achievement after one and two years of TIF implementation, potentially reflecting changes in individual students' achievement and changes in the schools' student composition resulting from pay-for-performance.[24]

### Association Between TIF Program Characteristics and Impacts (Chapter VI)

In Chapter VI (and Appendix G), we explore the association between districts' TIF program and implementation characteristics and the impacts of pay-for-performance. Districts varied substantially in the design and implementation of their TIF programs in ways that could have influenced the impacts of pay-for-performance. For example, some districts had pay-for-performance bonuses that were more differentiated for higher and lower performers than others, or had bonuses that were largely based on individual rather than group performance. Knowing whether the impacts of pay-for-

[21] Performance ratings in Year 1 could also reflect improvements in individual educators' performance resulting from pay-for-performance. Therefore, although the analysis described here may suggest the occurrence of staffing changes, it cannot definitively distinguish staffing changes from improvements by individual educators. For example, if performance ratings in Year 1 were higher among educators who stayed in treatment schools than among those who stayed in control schools, two potential explanations could be (1) pay-for-performance caused more higher-performing educators to stay or (2) pay-for-performance did not change who stayed or left, but motivated those who were intending to stay to work more effectively in Year 1.

[22] To ensure that all outcomes were measured in the spring, we used a grantee-administered test for one district located in a state with a fall state assessment.

[23] There were no differences between treatment and control schools in percentages of students in grades 3 through 8 who had math and reading scores in Years 1 and 2 (Appendix A, Table A.22). Compared to students without scores, those with scores were higher-achieving in the pre-implementation year, more likely to be female or Hispanic, and less likely to be black, have an Individualized Education Program, repeat a grade, or be overage for their grade (Appendix A, Tables A.23 and A.24).

[24] In Years 1 and 2, students in the analysis sample from treatment and control schools had similar characteristics, suggesting that pay-for-performance did not induce changes in the schools' student composition (Appendix A, Tables A.25 and A.26). Students from the analysis sample in treatment schools had lower baseline math achievement than students from the analysis sample in control schools, but this pattern simply mirrored the treatment-control difference in math achievement that we observed among students enrolled in the pre-implementation year (Table II.4).

performance were systematically larger or smaller in districts with particular program characteristics can suggest best practices for developing and improving these programs.

We assessed whether any of the following characteristics could help explain variation across districts in impacts on student achievement: (1) amount of differentiation in teachers' pay-for-performance bonuses, (2) districts' use of classroom achievement growth to measure teacher effectiveness, (3) the extent of teachers' understanding of their eligibility for pay-for-performance bonuses, and (4) the timing of the bonus award. We selected these four characteristics because of their potential to motivate teachers to change their behavior in response to pay-for-performance bonuses, which may, in turn, affect student achievement. For each feature, we categorized districts into two subgroups that differed according to the presence or absence of the characteristic, or according to whether districts had high or low levels of the characteristic. We then compared the impacts of pay-for-performance on student achievement in Year 2 between these two subgroups of districts. A significant difference in impacts between the two subgroups provides *suggestive* evidence that the characteristic may have influenced impacts, given that the two groups may differ on other measured and unmeasured characteristics.

THIS PAGE IS INTENTIONALLY BLANK

# III. PROGRAMS AND EXPERIENCES OF ALL 2010 TIF DISTRICTS

In this chapter, we broadly describe TIF program implementation in 2012–2013. We first examine how many TIF districts implemented all four required components of the TIF grant (discussed in Chapter I). We then provide more detail on the implementation of each individual component to examine which components contributed to districts' ability to implement (or inability to implement) all four required components. We conclude the chapter with details on challenges that districts reported encountering in implementing TIF.

The findings presented in this chapter are from 155 districts that were included in the 2010 TIF grants and implemented a TIF program in the 2012–2013 school year. The information in this chapter is based on surveys completed by TIF districts between June and August 2013, when nearly all TIF districts had completed their second year of program implementation. We also draw upon districts' responses to the 2012 surveys, which district staff completed approximately halfway through the first year of program implementation, to compare findings from the second year of implementation to those from the first year.

## TIF Required Components

The TIF grant required four components: (1) using student achievement growth and at least two formal observations to measure educator effectiveness, (2) offering a pay-for-performance bonus, (3) offering additional pay opportunities, and (4) providing professional development to support educators' understanding and use of the measures of effectiveness. Taken together, these components constitute a comprehensive performance-based compensation system.

**Key Findings on Programs and Experiences of All TIF Districts**

- **Full implementation of TIF continues to be a challenge, although districts' implementation from the first to the second year improved somewhat.**
- **Most districts implemented each individual required component of TIF, but were less likely to report offering targeted professional development and evaluating teachers and principals using both student achievement growth and at least two observations.**
- **Near the end of the second year of implementation, most districts reported that sustainability of their TIF program was a major challenge; however few reported other key activities related to their program were a major challenge.**

### Implementation of TIF Required Components

**Full implementation of TIF continues to be a challenge, although districts' implementation from the first to the second year improved somewhat.** Although 90 percent of all TIF districts in the second year (2012–2013) reported implementing at least 3 of the 4 required components for teachers, about one-half (52 percent) reported implementing all four (Table III.1). This was a slight improvement from the first year (2011–2012), when 85 percent of districts reported implementing at least 3 of the 4 required components and 46 percent reported implementing them all.

More than half of the districts (58 percent in Year 1 and 60 percent in Year 2) implemented all required components for principals aside from professional development.[25]

**Most districts implemented each of the four individual required components of TIF, but were least likely to report offering targeted professional development and evaluating teachers and principals using both student achievement growth and at least two observations.** In Year 2, nearly all the districts (over 90 percent) reported offering teachers and principals bonuses based on their performance and offering educators opportunities to earn additional pay (Table III.1). In contrast, approximately three-quarters of the districts reported that they offered the required professional development to their teachers, 80 percent reported using both student achievement growth and classroom observations to measure teacher effectiveness, and 65 percent reported using both student achievement growth and observations of school practices to measure principal effectiveness.

**Table III.1. TIF Districts' Reported Implementation of TIF Required Components for Teachers and Principals (Percentages)**

| | Year 1 (2011–2012) | Year 2 (2012–2013) |
|---|---|---|
| **Teachers** | | |
| Requirement 1: Measures of educator effectiveness[a] | 79 | 80 |
| Requirement 2: Pay-for-performance bonus | 94 | 98 |
| Requirement 3: Additional pay opportunities[b] | 86 | 91 |
| Requirement 4: Professional development | 66 | 74 |
| Implemented requirements 1, 2, and 3 | 68 | 71 |
| Implemented three of four requirements | 85 | 90 |
| Implemented all requirements | 46 | 52 |
| **Principals** | | |
| Requirement 1: Measures of educator effectiveness[a] | 68 | 65 |
| Requirement 2: Pay-for-performance bonus | 94 | 99 |
| Requirement 3: Additional pay opportunities[b] | 86 | 91 |
| Implemented requirements 1, 2, and 3[c] | 58 | 60 |
| **Number of Districts—Range[d]** | **137–153** | **142-155** |

Source: Max et al. (2014); district survey, 2013.

[a]TIF districts were required to use student achievement growth and at least two observations by trained observers to evaluate teachers and principals.

[b]The TIF grant notice required that districts provide additional pay opportunities for educators, so these percentages are based on the percentage of TIF districts that reported they offered these pay opportunities to either teachers or principals.

[c]The district survey did not include questions on professional development for principals.

[d]Sample sizes are presented as a range based on the data available for each row in the table.

[25] Professional development for principals is a requirement of TIF grants. However, given concerns about the length of the district survey, it did not include questions on whether districts implemented the required professional development for principals. The TIF notice also required pay for additional opportunities for educators. Most grantees met this requirement by offering additional pay opportunities to teachers. Therefore, if the district reported offering additional pay opportunities to either teachers or principals, they met this requirement.

Next, we provide an overview of districts' implementation of each individual required component in 2012–2013.[26]

**Requirement 1: Measures of Educator Effectiveness**

TIF grantees were required to measure educator effectiveness based on student achievement growth and multiple observations by trained observers. These measures provide the basis for teachers and principals earning performance-based bonuses.

**Most TIF districts reported meeting the requirement to use student achievement growth and at least two observations to measure teacher and principal effectiveness.** Eighty percent of TIF districts reported using student achievement growth and classroom observations to measure teacher effectiveness, and 65 percent reported meeting the requirement to measure principal effectiveness (Table III.1).

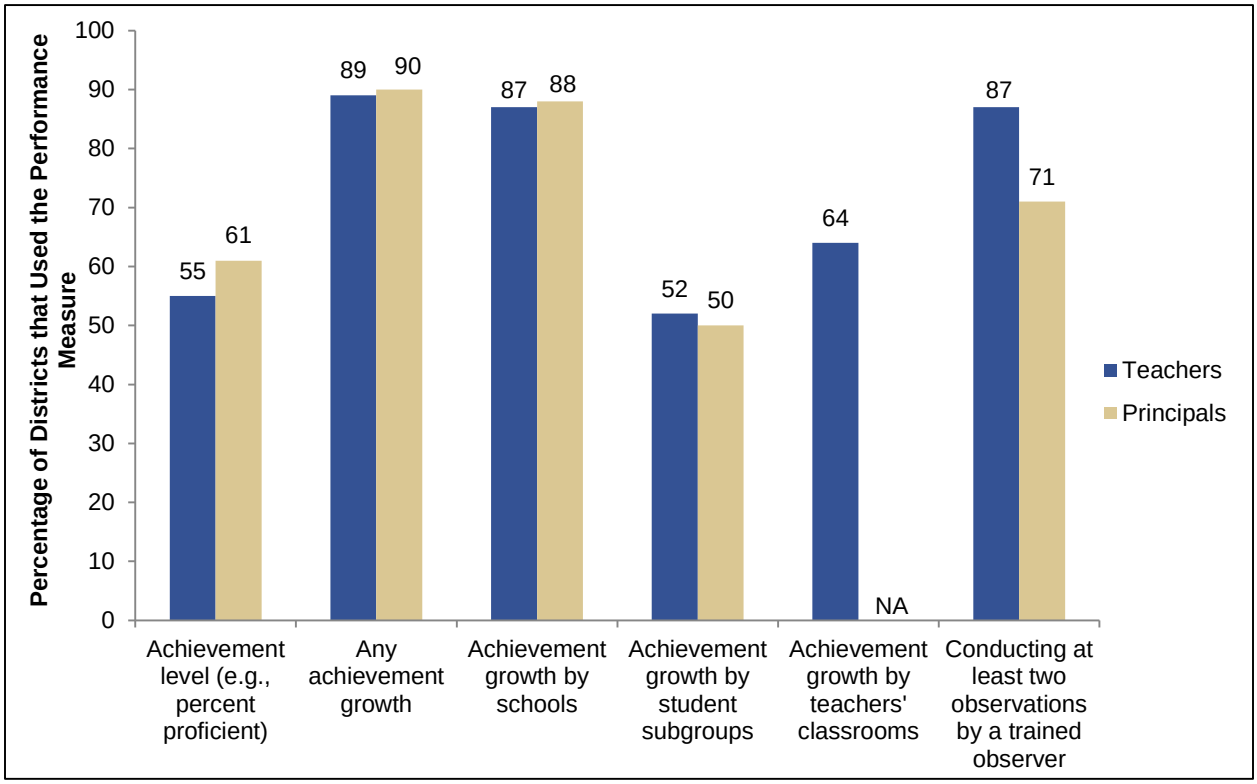
**Nearly all TIF districts reported using school-level student achievement growth to evaluate teachers.** Districts could choose whether to evaluate teachers based on achievement growth in their own classrooms, achievement growth for the entire school, achievement growth for a subgroup (such as an entire grade level), or a combination of these measures (Figure III.1). Classroom achievement growth measures could give teachers more control over their own evaluation ratings, and achievement growth measures for larger groups could encourage collaboration among teachers. Most frequently, TIF districts reported using school achievement growth measures (87 percent), followed by classroom achievement growth measures (64 percent) and measures of achievement growth by student subgroups (52 percent).

**Most TIF districts reported using at least two formal observations to evaluate teachers.** Eighty-seven percent of districts reported using at least two formal observations by trained observers to evaluate teachers (Figure III.1). Districts planned to conduct, on average, 3.5 formal observations per teacher—more than the two required under the grant—lasting about 45 minutes each (Appendix C, Table C.1). Districts most frequently reported that observations were conducted by principals (97 percent).

**Most TIF districts reported using student achievement growth and observations by trained observers to evaluate principals.** Most frequently, districts reported using school achievement growth to evaluate principals (88 percent) (Figure III.1). Most districts (71 percent) also reported conducting observations by trained observers. Districts planned to conduct, on average, about three observations per principal, lasting nearly an hour (54 minutes) each (Appendix C, Table C.1). Districts most frequently reported that observations of principals were conducted by a central office administrator from the same district (58 percent).

[26] Districts' implementation of each required component in 2012–2013 was similar to their implementation of each component in 2011–2012 (Max et al. 2014).

**Figure III.1. Measures of Student Achievement and Observations Used to Evaluate Teachers and Principals, All TIF Districts, Year 2 (Percentages)**




Source: District survey, 2013.

Note: Between 148 and 151 districts responded to the survey questions for teachers, and between 150 and 153 districts responded to the survey questions for principals.

Figure reads: In Year 2, 87 percent of all TIF districts reported using achievement growth by schools to evaluate teachers, and 88 percent reported using achievement growth by schools to evaluate principals.

NA is not applicable.

## Requirement 2: Pay-for-Performance Bonuses

TIF districts were required to offer pay-for-performance bonuses to teachers and principals based purely on their performance, but districts could determine which types of teachers would be eligible for such bonuses and whether other school staff would also be eligible. The determination of who is eligible could affect educators' attitudes toward and responses to their TIF programs. For example, broadening eligibility for bonuses to all staff at a school might increase the staff's buy-in to the program and, if bonuses depend on school performance measures, encourage collaboration among staff. Alternatively, limiting eligibility to teachers of certain grades or subjects might enable districts to concentrate resources on improving classroom practices in high-priority academic areas.

**Most TIF districts sought to make performance bonuses broadly available to a variety of school staff.** Nearly all TIF districts reported that teachers and principals were eligible for pay-for-performance bonuses. Ninety-seven percent of TIF districts reported that teachers were eligible for performance bonuses, and 99 percent reported that principals were eligible (Table III.2). Teachers' eligibility for performance bonuses was almost never contingent upon teaching a grade or subject with annual, end-of-year state assessments. In fact, 97 percent of districts reported that teachers in grades or subjects without annual assessments (referred to as "nontested") were eligible for performance bonuses (Table III.2). Moreover, districts tended not to restrict eligibility to teachers and principals.

Eighty-two percent of districts reported that assistant/vice principals were eligible for performance bonuses. Almost half of districts (48 percent) reported making nonteaching staff, such as counselors, librarians, or custodians, eligible for such bonuses.

**Table III.2. Staff Eligibility for Pay-for-Performance Bonus, Year 2 (Percentages)**

| | All TIF Districts |
|---|---|
| Teachers | |
| Teachers in tested grades and subjects | 97 |
| Teachers in nontested grades and subjects | 97 |
| Principals | 99 |
| Other School Staff | |
| Assistant/vice principal | 82 |
| Other school administrators | 30 |
| Other teaching staff (e.g., part-time teachers, substitutes, aides) | 20 |
| Nonteaching staff (e.g., counselors, librarians, custodians) | 48 |
| **Number of Districts—Range[a]** | **126-155** |

Source: District survey, 2013.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

### Requirement 3: Additional Pay Opportunities

TIF programs had to include financial incentives for educators to take on additional roles and responsibilities. Examples from the TIF notice included serving as a master or mentor teacher whose roles typically include mentoring novice teachers, developing professional learning communities, and tutoring students. By identifying highly effective teachers and encouraging these teachers to share their expertise with their colleagues, TIF grants could improve overall teacher effectiveness in high-need schools. These types of additional pay opportunities may also help attract or retain highly effective teachers who seek these leadership roles.

**Nearly all TIF districts offered additional pay for teachers to take on roles and responsibilities, most often to support mentor or master/lead teacher opportunities.** Ninety percent of TIF districts reported offering teachers additional pay for roles and responsibilities (Table III.3). Most often, districts offered additional pay for mentor (64 percent) and master or lead teachers (62 percent). Few districts (13 percent) reported offering principals extra pay for assuming additional roles or responsibilities.

The TIF notice also encouraged, but did not require, districts to offer additional pay for educators to teach in high-need subject areas or to work in hard-to-staff schools. A minority of districts (39 percent) offered teachers additional pay for doing so (Appendix C, Table C.2).

**Table III.3. Additional Pay Opportunities for Teachers and Principals, Year 2**

| | Percentage of TIF Districts That Offered Additional Pay | Average Maximum Pay in Districts Offering Additional Pay |
|---|---|---|
| **Teachers** | | |
| Teachers could receive additional pay for taking on extra roles or responsibilities. | 90 | NA |
| Roles and responsibilities | | |
| Mentor teacher | 64 | $4,878 |
| Master or lead teacher | 62 | $7,841 |
| Department chair or head | 23 | $2,100 |
| Lead curriculum specialist | 14 | $2,965 |
| Schoolwide committee or task force member | 22 | $798 |
| Leadership team member | 18 | $1,904 |
| **Number of Districts—Range[a]** | **123–154** | **20–137** |
| **Principals** | | |
| Principals could receive additional pay for taking on extra roles or responsibilities in school or district. | 13 | $3,258 |
| **Number of Districts—Range[a]** | **123-155** | **18-94** |

Source: District survey, 2013.

Note: Table reports on activities funded by TIF.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

NA is not applicable.

## Requirement 4: Professional Development

The TIF notice required that districts provide professional development linked to the measures of educator effectiveness. This support included professional development to help educators understand the measures being used to evaluate their performance, as well as feedback based on their actual performance ratings to help improve their instructional practices.

**Approximately three-quarters of the TIF districts provided the required professional development to teachers.** Although nearly all TIF districts (94 percent) offered professional development to help teachers understand the performance measures used in the program, fewer districts (76 percent) offered the more targeted professional development based on teachers' actual performance (Table III.4).[27]

**Table III.4. Planned Professional Development Activities for Teachers, Year 2 (Percentages)**

| | All TIF Districts |
|---|---|
| Understanding performance measures of TIF program | 94 |
| Feedback based on TIF performance ratings | 76 |
| **Number of Districts** | **152** |

Source: District survey, 2013.

[27] Surveys of district administrators did not ask about professional development for principals.

## Challenges in Implementing and Sustaining TIF

In addition to asking about implementation of each required component, the 2013 district survey included questions about challenges districts faced implementing TIF. Our goal was to focus on topics that might shed light on the components that could make it difficult for districts to implement programs like TIF. The survey asked district staff whether particular aspects of implementation were a "major challenge," "minor challenge," or "not a challenge." For example, we asked about potential challenges related to (1) incorporating student achievement growth into teacher evaluations, (2) observing teachers' or principals' practices, (3) calculating pay-for-performance bonuses, (4) communicating the program to educators or other stakeholders, (5) obtaining or maintaining support for the program, and (6) sustaining the program. This section focuses on the activities that districts most often reported as a major challenge.[28]

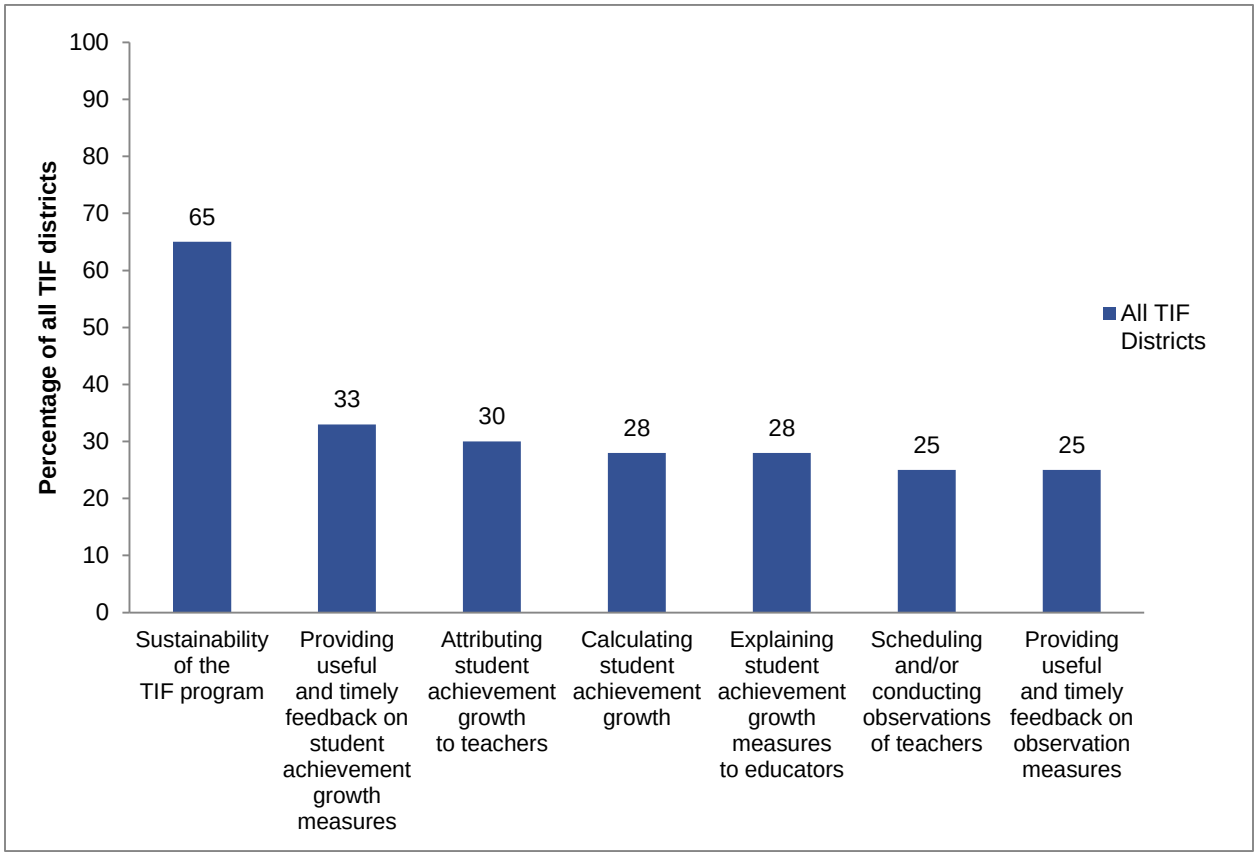
**Near the end of the second year of implementation, most districts reported that sustainability of their TIF program was a major challenge; however, few reported other key activities related to their program were a major challenge.** Applicants for a TIF grant had to provide evidence that they could sustain their performance-based compensation systems beyond the life of their TIF grant, and grantees were required to fund an increasing share of the program costs over the course of the five-year grant. By the end of 2012–2013 (which, for most districts, was the second of four years of implementation under the grant), 65 percent of TIF districts reported that sustainability of the program was a major challenge (Figure III.2). In contrast, fewer than one-third of districts reported that linking student growth data to teachers (30 percent), explaining student achievement growth to teachers (28 percent), and calculating student achievement growth to evaluate teachers (28 percent) were major challenges. Likewise, only one-third of districts (33 percent) reported that providing useful and timely feedback on student achievement measures was a major challenge.

One way for districts to address challenges is to revise their program. Because of the potential for pay-for-performance to be challenging for districts, the district survey focused on what revisions, if any, districts reported making to their performance bonuses after the first year of implementation.

**Most TIF districts reported that they did not make revisions to their performance bonuses after the first year.** Consistent with most districts' reports that implementing pay-for-performance bonuses was not a major challenge, only thirty-five percent of districts reported making a change after the 2011–2012 school year to some aspect of their program related to pay-for-performance bonuses (Appendix C, Table C.4). These revisions included changing the evaluation criteria for earning a performance bonus (32 percent) and expanding eligibility for pay-for-performance bonuses (13 percent). The most commonly reported reasons for revising pay-for-performance bonuses were to improve the perceived fairness of the bonuses (20 percent) and to obtain or maintain support from stakeholders (14 percent) (Appendix C, Table C.5).

[28] Appendix C, Table C.3 shows a full list of activities included in the survey and the percentages of districts that reported these activities to be a major challenge, minor challenge, and not a challenge. Since this was the first time the district survey asked about challenges, the questions asked generally if districts found these issues challenging to implement. The questions did not specify if they currently found these issues challenging.

**Figure III.2. Major Challenges in Implementing TIF, Year 2 (Percentages)**




Source: District survey, 2013.

Note: Between 147 and 155 TIF districts responded to these survey questions. Further details about survey results, including results for activities that districts reported as a "minor challenge" or "not a challenge," can be found in Appendix C, Table C.3.

Figure reads: In Year 2, 65 percent of all TIF districts reported that sustainability of their TIF program was a major challenge and 25 percent reported that providing useful and timely feedback on observation measures was a major challenge.

## Summary

As a comprehensive program for reforming educator compensation and improving educator effectiveness, TIF programs were designed to have multiple, interrelated components. Our analysis of implementation in all 155 TIF districts sought to determine whether they could put into place such a comprehensive system, and whether they faced particular challenges doing so.

Overall, the 2010 TIF districts were able to implement most required components of a comprehensive performance-based compensation system – without major, widespread challenges. Although there was some improvement in districts' implementation by the second year, many districts still did not implement all the required components. Providing professional development to help educators understand how they were being evaluated was the districts' most common reason for not achieving full implementation of TIF for teachers. By the third year of the grant, most TIF districts believed that sustaining their system will be a major challenge.

# IV. TIF IMPLEMENTATION IN EVALUATION DISTRICTS

In this chapter, we describe the implementation of TIF by the evaluation districts—those that were awarded a grant to participate in the evaluation of TIF, including random assignment of the pay-for-performance component of the program. According to the theory of change presented in Chapter I, a series of steps needed to occur in the implementation of TIF for pay-for-performance to be able to improve educator effectiveness and student achievement. The components of the program needed to provide incentives and supports for educators to improve their effectiveness, information about those components needed to be communicated to educators, and educators needed to receive and understand this information. This chapter examines whether and how each of these steps materialized in the evaluation districts' implementation of TIF. First, we examine districts' implementation of the four required components of TIF. We focus on aspects of the programs that could shape teachers' motivation to improve (such as whether performance measures provided educators with consistent information on their effectiveness and whether pay-for-performance bonuses were differentiated, substantial, and challenging to earn[29]). Second, we examine how districts communicated information about TIF to educators, including information on the specific bonus amounts that educators received. In the final part of this chapter, we examine teachers' and principals' understanding of the TIF program in their districts. Describing the implementation of the TIF grant in evaluation districts is useful context for interpreting findings presented later in this report on the program's impact on educator and student outcomes.

The chapter is based on 10 evaluation districts that completed two years of TIF implementation during the period covered by this report. We refer to the first and second years of implementation, 2011–2012 and 2012– 2013, as Years 1 and 2.[30] In both years, educators in treatment schools were eligible for pay-for-performance bonuses, and educators in control schools were not. The information in this chapter is drawn from details we obtained from these districts through district, teacher, and principal surveys; interviews with district TIF administrators; administrative data provided by the districts; and technical assistance documents.

[29] The TIF grant notice provided examples of bonuses that were differentiated, substantial, and challenging to earn. We describe these examples later in this chapter, as well as in Chapter I.

[30] As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts examined in this chapter, whose schools were randomly assigned in spring and summer 2011, were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort. In Appendix D, we present key implementation findings from Year 1 for Cohorts 1 and 2 combined—that is, findings from 2011–2012 for Cohort 1 and 2012–2013 for Cohort 2.

**Key Findings on TIF Implementation in Evaluation Districts**

- **Most districts reported implementing all required components for teachers, and all districts reported meeting at least three of the four required components for teachers and principals. The only component not consistently implemented was professional development for teachers.**
- **All evaluation districts reported using the achievement growth of all students in a school to evaluate teachers, and some also chose to evaluate teachers based on the achievement growth of the students they teach.**
- **Student achievement growth and observation ratings sometimes identified the same educators as high-performing, but many earned higher ratings on observations than on achievement growth.**
- **At least half of districts met the TIF grant guidance for awarding differentiated pay-for-performance bonuses for teachers, but not the guidance for awarding bonuses that were substantial or challenging to earn.**
- **Most districts' performance bonuses for principals were not differentiated, substantial, or challenging to earn.**
- **Teachers' understanding of performance measures improved between the first and second years of implementation, as did teachers' and principals' understanding of their eligibility for pay-for-performance bonuses.**
- **Many teachers in schools that offered pay-for-performance bonuses still did not understand that they were eligible for a bonus or underestimated how much they could earn from performance bonuses.**

## Implementation of the Required Components of TIF

Our examination of the implementation of TIF programs in evaluation districts focuses on the four required components of TIF programs: (1) measures of educator effectiveness, (2) pay-for-performance bonuses, (3) additional pay opportunities, and (4) professional development. Together, these four required components constitute a comprehensive performance-based compensation system, and the grant required that all the individual components be implemented together. In this section, we report on TIF evaluation districts' success in implementing all components together and on their implementation of each component separately.

### Implementation of All Required Components

**Most evaluation districts reported implementing all required components for teachers, and all districts reported meeting at least three of the four required components. The only component not consistently implemented was professional development.** Seventy percent of evaluation districts implemented all four required components for teachers. All evaluation districts reported using a measure of effectiveness that included students' achievement growth and at least two observations of classroom practices, offering bonuses based on how teachers performed on effectiveness measures, and offering additional pay to take on extra roles or responsibilities (Table IV.1). However, only 7 of 10 evaluation districts reported providing the required professional

development (similar to all 2010 TIF districts). The percentage of districts meeting each requirement was identical in Years 1 and 2.

**Table IV.1. Evaluation Districts' Reported Implementation of TIF Program Requirements for Teachers and Principals (Percentages)**

| | Year 1 (2011–2012) | Year 2 (2012–2013) |
|---|---|---|
| **Teachers** | | |
| Requirement 1: Measures of educator effectiveness[a] | 100 | 100 |
| Requirement 2: Pay-for-performance bonus | 100 | 100 |
| Requirement 3: Additional pay opportunities[b] | 100 | 100 |
| Requirement 4: Professional development | 70 | 70 |
| Implemented requirements 1, 2, and 3 | 100 | 100 |
| Implemented all requirements | 70 | 70 |
| **Principals** | | |
| Requirement 1: Measures of educator effectiveness[a] | 70 | 100 |
| Requirement 2: Pay-for-performance bonus | 100 | 100 |
| Requirement 3: Additional pay opportunities[b] | 100 | 100 |
| Implemented requirements 1, 2, and 3[c] | 70 | 100 |
| **Number of Districts** | **10** | **10** |

Source: District surveys (2012 and 2013) and district interviews, 2012 and 2013.

[a]TIF districts were required to use student achievement growth and at least two observations by trained observers to evaluate teachers and principals.

[b]The TIF grant notice required that districts provide additional pay opportunities for educators, so these percentages are based on the percentage of TIF districts that reported they offered these pay opportunities to either teachers or principals.

[c]We do not have data on the percentage of districts that provided professional development to principals.

**All evaluation districts also reported meeting three of the four required components for principals.** In addition, evaluation districts made progress between Years 1 and 2 in implementing the required components for principals. Although only 70 percent of districts implemented the required measures of effectiveness for principals in Year 1, all districts did so in Year 2. In both years, all districts offered pay-for-performance bonuses to principals. Districts could meet the third requirement—additional pay opportunities—by providing opportunities to either teachers or principals; as discussed above, all districts fulfilled this requirement. We were unable to assess whether districts implemented the fourth required component for principals—professional development—because we did not have such data for principals.

Next, we describe implementation of each required component in more detail and compare the implementation between Years 1 and 2.

### Requirement 1: Measures of Educator Effectiveness

TIF grantees were required to measure educator effectiveness based on student achievement growth and multiple observations by trained observers. These measures provided the basis for rewarding teachers and principals with performance bonuses. As discussed earlier, by Year 2, all evaluation districts reported evaluating teachers and principals using the criteria required by the grant.

However, districts had discretion in designing the achievement growth and observation measures they used. Therefore, in what follows, we first describe the performance measures that districts reported using to evaluate teachers and principals. We then use administrative data to document teachers' actual performance on those measures, focusing on whether different measures were consistent with each other in assessing how well teachers performed. Although different performance measures may be designed to evaluate different aspects of performance, teachers may have trouble deciding whether and how to adjust their teaching practices if they receive conflicting information about their performance from different measures.

When districts designed their performance measures, one area of discretion involved choosing how to evaluate teachers based on student achievement growth. For example, districts could choose to evaluate teachers based on the achievement growth of the teachers' own students (classroom achievement growth), all students in the same grade, team, or subject area (achievement growth of student subgroups), all students in the school (school achievement growth), or some combination of these measures. Districts could measure student achievement growth using a value-added model or by calculating the change in students' achievement on a standardized test from one year to the next.

**All evaluation districts reported using school achievement growth to evaluate teachers, and some also chose to evaluate teachers based on the achievement growth of the students they teach.** To evaluate teachers in Year 2, all evaluation districts reported using school achievement growth, 60 percent reported using classroom achievement growth, and 30 percent reported using achievement growth of student subgroups (Table IV.2).

To evaluate principals, all evaluation districts used school achievement growth, and half used achievement growth of student subgroups (Table IV.2).

**Table IV.2. Measures of Student Achievement and Observations of Practices Used to Evaluate Teachers and Principals, as Reported by Districts, Year 2 (Percentages)**

| Performance Measure | Teachers | Principals |
|---|---|---|
| Student Achievement | | |
| Student achievement level (e.g., percent proficient) | 20 | 50 |
| Student achievement growth | 100 | 100 |
| By school | 100 | 100 |
| By student subgroups[a] | 30 | 50 |
| By teacher's classroom | 60 | NA |
| Observation Measure | | |
| Conducting at least two observations by trained observer | 100 | 100 |
| **Number of Districts** | **10** | **10** |

Source: District survey, 2013.

[a]Examples of student subgroups include grouping students by grade, team, or subject area.

NA is not applicable.

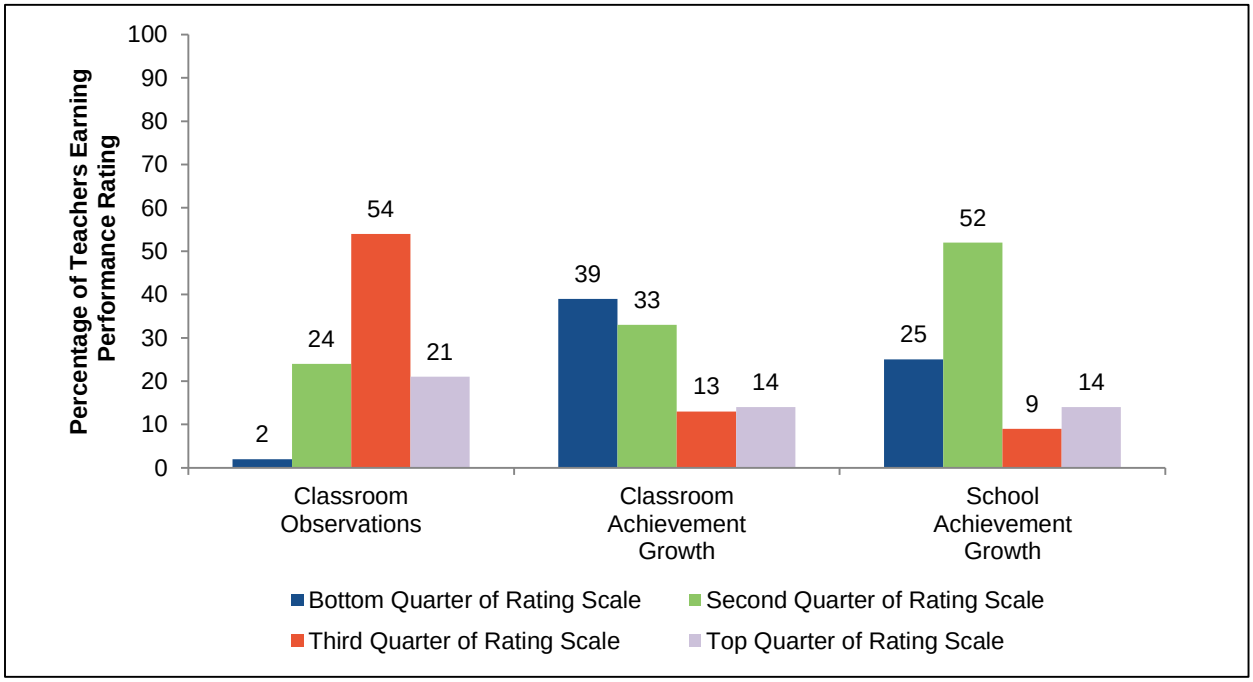
Among districts that used a particular type of achievement growth measure (such as school achievement growth), there were differences in how those measures were designed. For example, a review of technical assistance documents found that six evaluation districts used growth measures provided by the state, and four districts used models developed by private vendors.

Districts also had discretion in meeting the requirement to conduct observations of classroom or school practices. For example, districts could decide which rubrics they wanted to use to observe teachers and principals, the number of observations in a year (as long as there were at least two), and

which staff to train as observers. In practice, three districts used the Teacher Advancement Program (TAP) teacher observation rubric, three used Danielson's Framework for Teaching rubric (or a modified version of it), and two districts used a modified version of Kim Marshall's observation rubric. The remaining two districts used an existing state or district teacher observation rubric. On average, evaluation districts reported conducting 3.5 classroom observations per year, each about 50 minutes long. Most often, evaluation districts reported that classroom observations were conducted by the principal or other administrators at the teacher's school (90 percent), although half of the districts also reported that teacher leaders or peer observers conducted classroom observations (Appendix D, Table D.1).

**In the evaluation districts, observation ratings were typically at least moderately high even though student achievement growth ratings typically were not.** For example, in Year 2, 75 percent of teachers earned classroom observation ratings in the top half of the rating scale (54 percent in the third quarter and 21 percent in the top quarter of the rating scale), but only 23 percent of teachers earned school achievement growth ratings in the top half of the scale (Figure IV.1). Similar patterns were observed for teachers in Year 1 (Appendix D, Figure D.1)[31] and for principals in both years (Appendix D, Table D.2).

**Figure IV.1. Distribution of Teachers' Performance Ratings in Year 2**




Source: Educator administrative data (N = 3,628 teachers for the classroom observation score rating, N = 1,342 teachers for the classroom student achievement growth rating, and N = 4,432 teachers for the school student achievement growth rating).

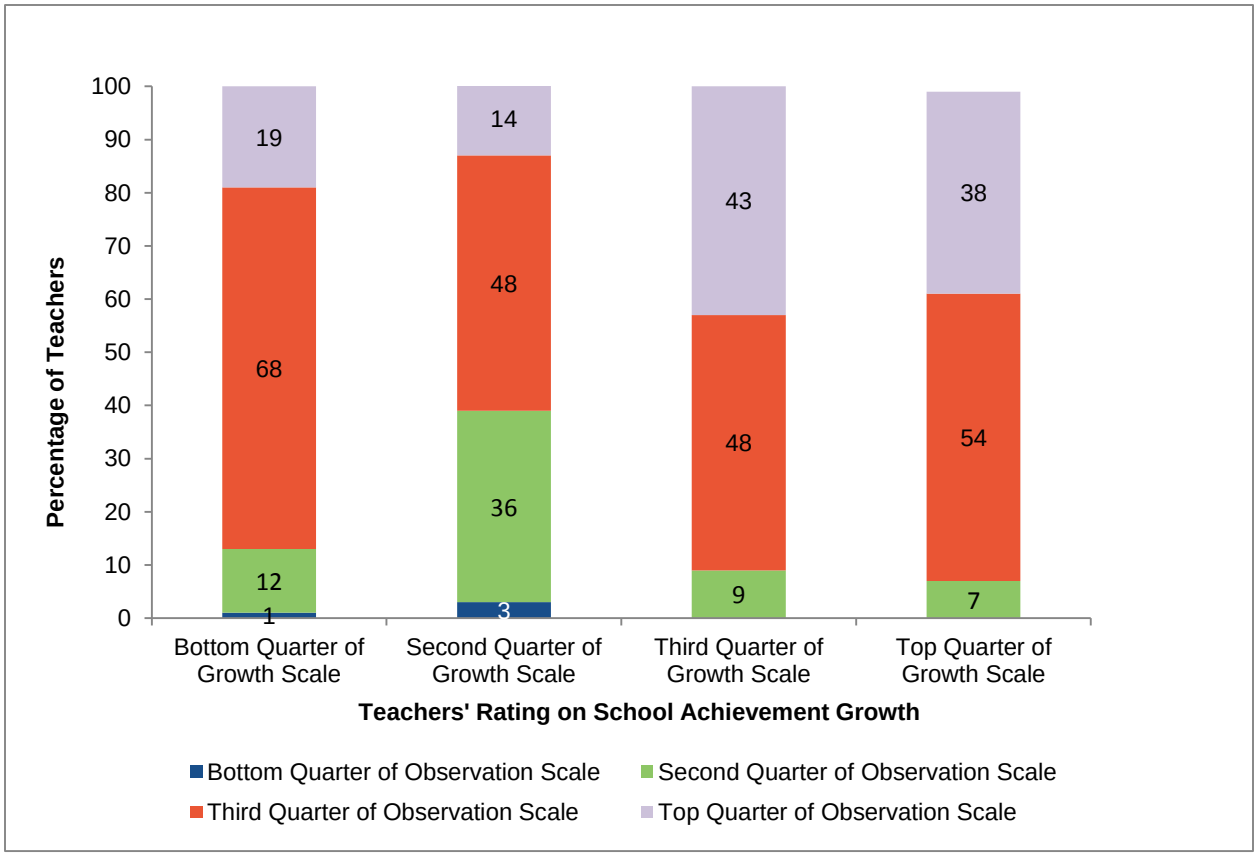
Note: On a 1 to 4 rating scale, the bottom quarter of the rating scale consists of ratings from 1 to 1.75, the second quarter ranges from 1.75 to 2.5, the third quarter ranges from 2.5 to 3.25, and the fourth quarter ranges from 3.25 to 4.

Figure reads: For classroom observations in Year 2, 2 percent of teachers were rated in the bottom quarter of the rating scale, 24 percent were rated in the second quarter, 54 percent in the third quarter, and 21 percent in the top quarter.

[31] We found a similar pattern for teachers in Year 1 based on both Cohorts 1 and 2 (Appendix D, Figure D.2).

**Student achievement growth and observation ratings sometimes identified the same educators as high-performing, but many earned higher ratings on observations than on achievement growth.** For example, in Year 2, teachers who scored high (in the top quarter of the rating scale) on school achievement growth were twice as likely to score high on classroom observations compared with teachers who scored low (in the bottom quarter of the rating scale) on school achievement growth (38 versus 19 percent; Figure IV.2). Nevertheless, many teachers (87 percent) who scored low on school achievement growth earned at least moderately high ratings on classroom observations, scoring in the top half of the observation rating scale. Likewise, many principals (78 percent) who scored low on school achievement growth in Year 2 earned at least moderately high ratings on observations (Appendix D, Table D.7).[32]

**Figure IV.2. Classroom Observation Ratings of Teachers Who Earned Lower and Higher Ratings on School Achievement Growth in Year 2 (Percentages)**




Source: Educator administrative data (N = 1,047 teachers in bottom quarter of growth scale, N = 1,753 teachers in second quarter of growth scale, N = 336 teachers in third quarter of growth scale, N = 465 teachers in top quarter of growth scale).

Note: On a 1 to 4 rating scale, the bottom quarter of the rating scale consists of ratings from 1 to 1.75, the second quarter ranges from 1.75 to 2.5, the third quarter ranges from 2.5 to 3.25, and the fourth quarter ranges from 3.25 to 4.

Figure reads: In Year 2, among teachers who were rated in the bottom quarter of the school achievement growth scale, 19 percent were rated in the top quarter of the observation rating scale, 68 percent were rated in the third quarter of the observation rating scale, 12 percent were rated in the second quarter of the observation rating scale, and 1 percent were rated in the bottom quarter of the observation rating scale.

[32] We found similar patterns in Year 1, except that educators with high achievement growth ratings were about equally likely to score high on observations as educators with low achievement growth ratings (Appendix D, Tables D.3, D.4, and D.6).

Classroom achievement growth ratings had fewer stark discrepancies with observation ratings, but observation ratings were still often higher. Within the six districts that used classroom achievement growth, about 40 percent of teachers (typically, those who taught grades and subjects in which annual state assessments were administered) received classroom achievement growth ratings (Appendix A, Table A.14). Among these teachers, only 4 percent who scored low on classroom achievement growth in Year 2 had high observation ratings (Appendix D, Table D.5). Still, many teachers (47 percent) who scored low on classroom achievement growth were still received at least moderately high observation ratings.

**Requirement 2: Pay-for-Performance Bonuses**

The evaluation design was based on random assignment of the pay-for-performance bonus component of the TIF program to some schools (the treatment schools) and not others (control schools). As discussed in Chapter I, although districts had discretion to specify the structure of performance bonuses, the TIF grant notice provided guidance to these districts by giving examples of bonuses that were *substantial* (with an average payout worth 5 percent of the average educator salary), *differentiated* (with at least some educators receiving a payout worth three times the average payout), and *challenging* to earn (with only those performing significantly better than the average receiving bonuses). To describe these bonuses and their alignment with TIF grant guidance, we used administrative data provided by the TIF evaluation districts.

When designing performance bonuses, districts faced the key decision of whether to offer separate bonuses for different performance measures or combine all of the performance measures into a single rating that determined educators' bonuses. Awarding separate bonuses for different performance measures could make it easier for educators to understand why they did or did not receive a bonus. However, it also had the potential to make earning a bonus less challenging, because educators would need to perform well on only one measure to earn a bonus. Educators might even choose to focus improving their performance only on the measure (or measures) that they believed they could change most easily.

**All evaluation districts met the TIF grant requirement to offer teachers pay-for-performance bonuses, and all chose to offer separate bonuses for different performance measures.** In Year 2, all evaluation districts offered teachers bonuses based on school achievement growth, 70 percent of districts offered bonuses for classroom observations, 60 percent offered bonuses for classroom achievement growth, and 30 percent provided bonuses for achievement growth of student subgroups.[33] Most districts set an absolute maximum bonus that could be earned for each measure, but in some districts, the maximum bonus that could be earned depended on the number of bonus recipients (Table IV.3).[34]

[33] In contrast, most (67 percent) of the Cohort 2 districts used a single, combined performance rating to determine bonuses.

[34] Appendix D, Tables D.8 and D.9 provide summary and detailed information, respectively, on teacher pay-for-performance programs for Cohorts 1 and 2.

**Table IV.3. Key Features of Evaluation Districts' Teacher Pay-for-Performance Bonus Programs in Year 2**

| | Districts | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Key Program Feature | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Teachers could receive a bonus for multiple performance measures | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teachers could receive a bonus for school achievement growth | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teachers in tested grades and subjects could receive a bonus for their students' achievement growth | | | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ |
| Teachers could receive a bonus for the achievement growth of a student subgroup | | | | | ✓ | ✓ | | | ✓ | |
| Student achievement growth was measured by a value-added model | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| Teachers could receive a bonus for classroom observations | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | ✓ |
| A maximum bonus was specified for each performance measure | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Maximum bonus possible depended on the number of bonus recipients | ✓ | | ✓ | ✓ | | | | | | ✓ |
| Bonus amount for a performance measure could be affected by a factor besides the teacher's rating on the measure | | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| District changed some aspect of its program between Year 1 and Year 2 | ✓ | ✓ | | | | | ✓ | | ✓ | |

Source: District interviews from 2012 and 2013, grantees' Annual Performance Report (APR) documents, and technical assistance documents.

Note: Grantees submit an APR to the U.S. Department of Education that describes how educators are evaluated. To ensure district confidentiality, the numbers assigned to districts in Table IV.3 do not correspond to the letters assigned to districts in other parts of the report.

**For each type of measure, the approach to determine bonus amounts varied across districts and often included multiple criteria.** Most districts used a statistical model known as a value-added model to assess student achievement growth, but the criteria for earning a bonus still varied. For example, the bonus may have been based on the school's value-added score compared to the state mean, or compared to its own expected performance based on the prior year's performance, or based on its percentile ranking. In most districts, the bonuses that teachers received for a particular performance measure depended on additional factors beyond just their rating on that measure. For example, teachers' bonuses could vary depending on their attendance rate, whether they taught tested or untested grades, or whether they were career or mentor teachers. Likewise, three districts either reduced or withheld bonuses based on achievement growth from teachers who earned low classroom observation scores (Table IV.3).

**At least half of the districts met the TIF guidance to award differentiated performance bonuses for teachers.** Seventy percent of evaluation districts met the guidance for awarding differentiated performance bonuses for teachers in Year 1, and 50 percent met this guidance in Year

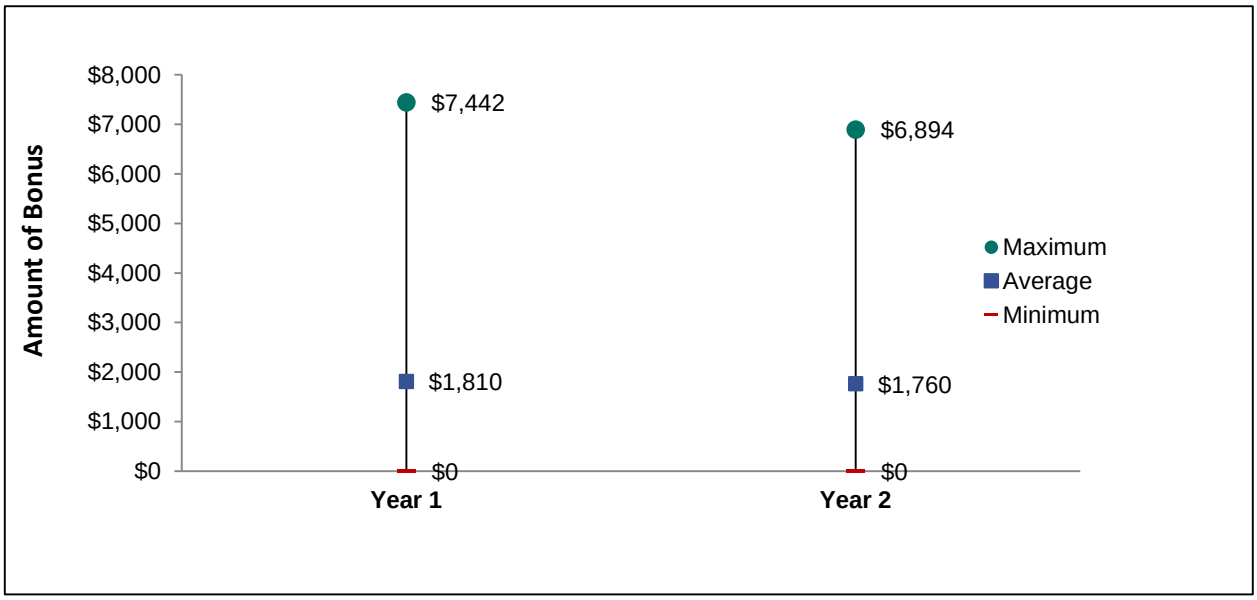
2 (Table IV.4).[35] On average across evaluation districts, the maximum bonus ($7,442 in Year 1 and $6,894 in Year 2) was more than three times the average bonus ($1,810 in Year 1 and $1,760 in Year 2) in treatment schools (Figure IV.3).[36]

**Table IV.4. Evaluation Districts Meeting TIF Grant Goals for Pay-for-Performance Bonuses for Teachers (Percentages)**

| TIF Grant Goal | Year 1 | Year 2 |
|---|---|---|
| Substantial: Average bonus was at least 5 percent of average salary | 20 | 20 |
| Differentiated: Highest bonus was at least three times the average bonus | 70 | 50 |
| Challenging: Less than 50 percent of teachers received a pay-for-performance bonus | 20 | 30 |
| **Number of Districts** | **10** | **10** |

Source: Educator administrative data.

**Figure IV.3. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers**




Source: Educator administrative data (N = 2,189 in Year 1 and N = 2,207 in Year 2).

Note: The statistics shown in this figure represent an equal-weighted average of the statistics from the 10 evaluation districts in Cohort 1. Year 1 findings were similar when districts were weighted by the number of schools (Appendix D, Figure D.3).

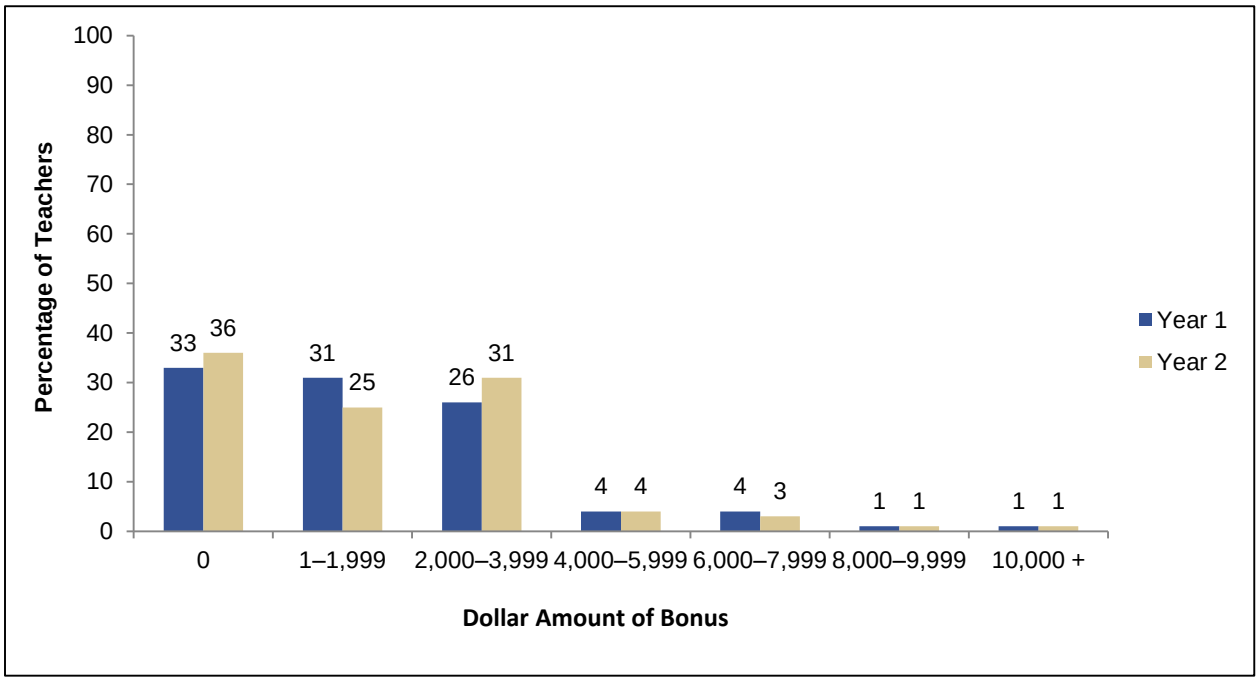
Figure reads: In Year 2, on average across the evaluation districts, the minimum pay-for-performance bonus was $0, the average pay-for-performance bonus was $1,760, and the maximum pay-for-performance bonus was $6,894.

[35] In Year 1, when findings were based on both Cohorts 1 and 2, 15 percent of districts met the guidance for awarding substantial bonuses, 69 percent met the guidance for awarding differentiated bonuses, and 31 percent met the guidance for awarding challenging bonuses (Appendix D, Table D.10).

[36] When Year 1 findings were based on both Cohorts 1 and 2, the average ($1,515) and maximum ($6,842) performance bonus amounts were slightly lower than the corresponding statistics for Cohort 1 only (Appendix D, Figure D.4).

**Few evaluation districts met the TIF guidance for awarding substantial and challenging performance bonuses for teachers.** Twenty percent of evaluation districts met the guidance for awarding substantial bonuses for teachers (Table IV.4). Across evaluation districts, the average bonus for treatment teachers was about $1,800, or about 4 percent of the average district salary (Figure IV.3).[37] In addition, fewer than one-third of the districts met the guidance for challenging bonuses (Table IV.4). Across districts, on average, more than 60 percent of treatment teachers received a bonus (Figure IV.4).[38]

**Figure IV.4. Distribution of Pay-for-Performance Bonuses for Teachers**




Source: Educator administrative data (N = 2,189 teachers in Year 1 and N = 2,207 teachers in Year 2).

Figure reads: In Year 2, 36 percent of teachers did not receive a pay-for-performance bonus, and 25 percent received a pay-for-performance bonus between $1 and $1,999.

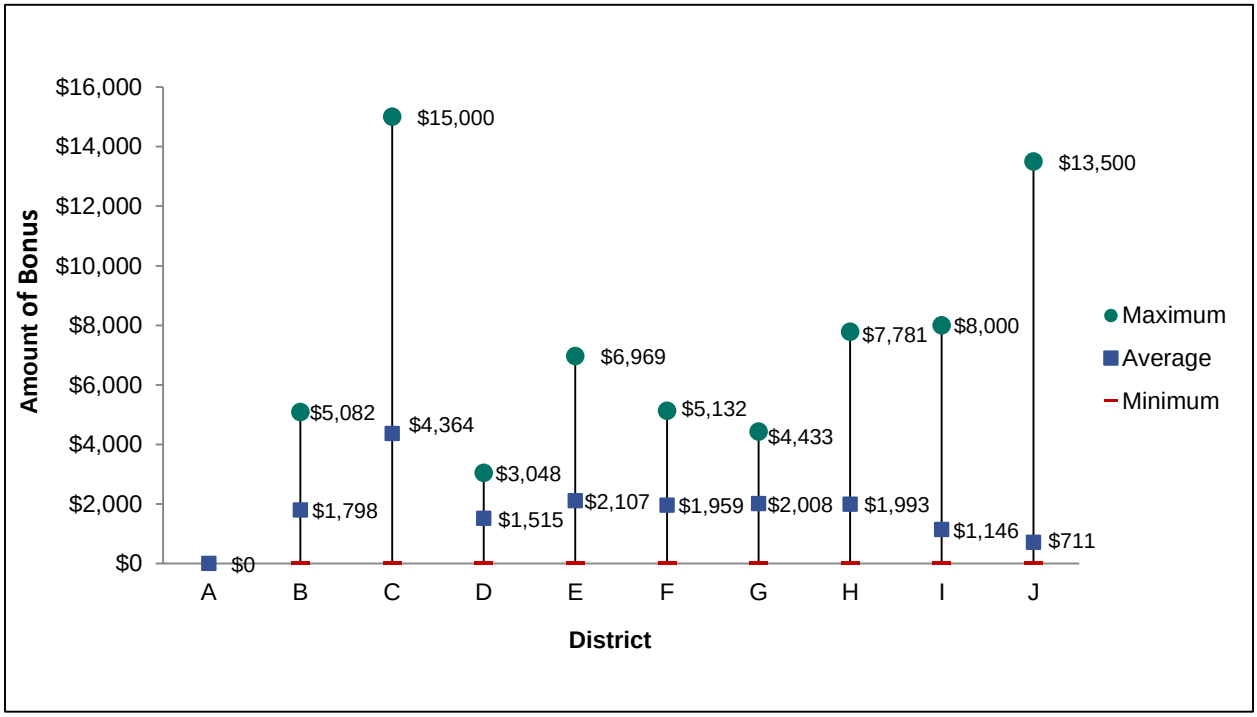
The percentage of teachers earning a bonus was one of several aspects of the bonus distributions that remained relatively stable across years. Sixty-seven percent of treatment teachers in Year 1 and 64 percent in Year 2 earned a bonus. Likewise, the percentages of teachers who earned bonus amounts in specific ranges were similar across years. For example, 26 percent of treatment teachers received a performance bonus between $2,000–3,999 in Year 1, and 31 percent of treatment teachers received performance bonuses in this range in Year 2 (Figure IV.4).

[37] We calculated whether bonuses were substantial using the average teacher salary that districts specified during interviews. The unweighted average salary across the 10 evaluation districts was about $48,000 for teachers and $89,000 for principals.

[38] Appendix D, Figures D.6 and D.7 show the percentage of teachers who earned a bonus, by district. Most districts in Year 1 (Cohorts 1 and 2) awarded performance bonuses to at least 50 percent of their treatment teachers. However, two districts awarded no performance bonuses, one district awarded bonuses to fewer than 1 percent, and one district awarded bonuses to 31 percent of its teachers (Appendix D, Figure D.6). We found similar patterns for Cohort 1 in Year 2 (Appendix D, Figure D.7).

**Maximum performance bonus amounts for teachers varied substantially across districts.** The average of the 10 districts' maximum performance bonus amounts (shown in Figure IV.3) masks considerable differences across districts in the maximum bonus that teachers earned. In Year 2, maximum performance bonus amounts were at least $13,500 in two districts, between $4,400 and $8,000 in six districts, and less than $3,100 in two districts (Figure IV.5).[39] Maximum performance bonus amounts varied to a similar extent in Year 1 (Appendix D, Figure D.5).[40] This variation suggests that setting the range of performance bonuses was an important dimension on which the evaluation districts chose to use their discretion in designing their TIF program and made substantially different choices.

**Figure IV.5. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers in Year 2, by District**




Source: Educator administrative data (N ranges from 81 teachers in District E to 394 in District J).

Figure reads: For District B in Year 2, the minimum pay-for-performance bonus was $0, the average pay-for-performance bonus was $1,798, and the maximum pay-for-performance bonus was $5,082.

[39] Due to circumstances beyond its control, one district (denoted by District A in Figure IV.5) was unable to distribute performance bonuses for Years 1 and 2 during the period covered by this report. However, educators were eligible for performance bonuses and were notified if they had earned a bonus and, if so, how much. For this reason, we treat District A as having met this requirement.

[40] There was also substantial variation in the maximum bonus amount among the Cohort 2 districts in Year 1, ranging from $0 to $8,525 (Appendix D, Figure D.8).

Because districts awarded separate bonuses for different performance measures, determining the amount of the bonus that was tied to each performance measure was a key decision that districts made to determine the structure of the incentives for teachers. Teachers could be motivated differently, depending on how much of their bonus was determined by measures of their individual performance (such as classroom observations and classroom achievement growth) as opposed to measures of team or school performance. On the one hand, tying larger bonuses to measures of individual performance could provide stronger motivation for individual teachers to change their teaching practices, because they have more control over their own performance than that of their team or school. On the other hand, tying larger bonuses to group performance measures might encourage collaboration.

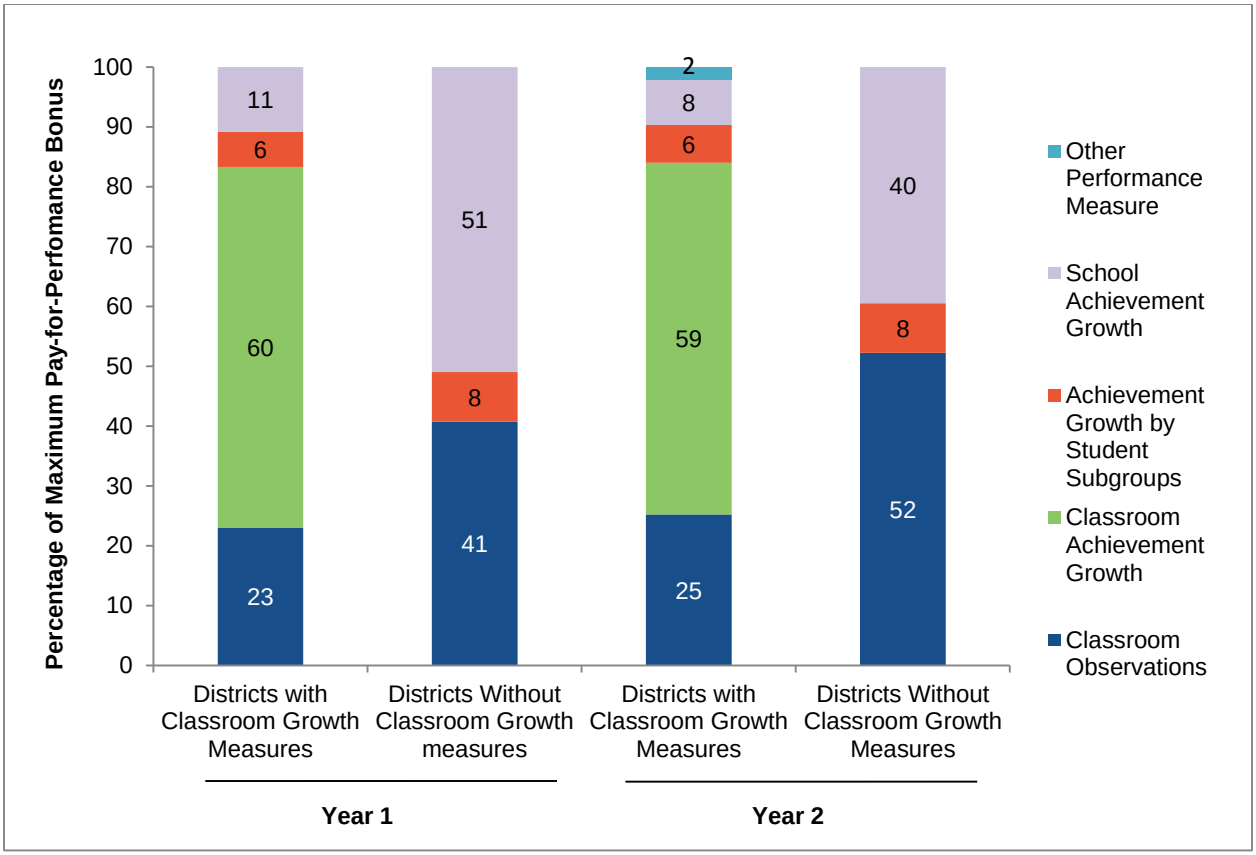
**In districts that used classroom achievement growth measures, a much larger share of the maximum bonus amount was based on teachers' individual performance rather than on the performance of groups of teachers.** In the six districts that had classroom achievement growth measures, the share of the maximum bonus tied to that classroom achievement growth was approximately 60 percent, the share tied to classroom observations was about 25 percent, and school achievement growth measures accounted for 11 percent or less of the maximum bonus. Therefore, in those districts, measures of individual performance—classroom achievement growth and classroom observations—together were the basis for most (nearly 85 percent) of the maximum bonus. In contrast, in the districts that did not use classroom achievement growth measures for performance bonuses, approximately half of the maximum bonus was based on classroom observation measures and half on school achievement growth measures (Figure IV.6).

In districts that used classroom achievement growth measures, teachers who were evaluated on those measures earned larger maximum bonuses than those who were not. Teachers who were evaluated on classroom achievement growth earned maximum performance bonuses of at least $7,600, and teachers who were not evaluated on those measures earned maximum bonuses of less than $4,800 (Appendix D, Figure D.9).

**All evaluation districts offered principals pay-for-performance bonuses, but few evaluation districts met the TIF guidance for awarding substantial, differentiated, or challenging pay-for-performance bonuses for principals.** All evaluation districts provided principals the opportunity to earn a bonus based on school achievement growth, and 9 of 10 offered principals bonuses based on at least one other performance measure, such as an observation rating or the achievement growth of student subgroups. Thirty percent of evaluation districts met the guidance for substantial bonuses for principals (Table IV.5). Across districts, the average bonus for treatment principals ($3,216 in Year 1 and $3,530 in Year 2) was no more than 4 percent of the average principal salary (Figure IV.6). Ten percent of districts met the guidance for differentiated bonuses in the first and second years (Table IV.5). On average across districts, the maximum bonus ($6,711 in Year 1 and $6,988 in Year 2) was only about twice the average bonus (Figure IV.7).[41] Twenty percent of the districts met the guidance for challenging bonuses in Year 1 or Year 2; at least 70 percent of principals in each year received a bonus (Appendix D, Figure D.12).

[41] When Year 1 findings were based on both Cohorts 1 and 2, the average ($2,690) and maximum ($5,378) performance bonus amounts were lower than the corresponding amounts for Cohort 1 only (Appendix D, Figure D.11).

**Figure IV.6. Teachers' Maximum Pay-for-Performance Bonus Attributable to Each Performance Measure (Percentages)**




Source: Educator administrative data (N = 1,004 teachers in Year 1 and N = 1,124 teachers in Year 2).

Note: Analyses are based on teachers who were evaluated on all major performance measures used by their districts. The figure excludes one district that did not make payouts. Examples of student subgroups include grouping students by grade, team, or subject area.

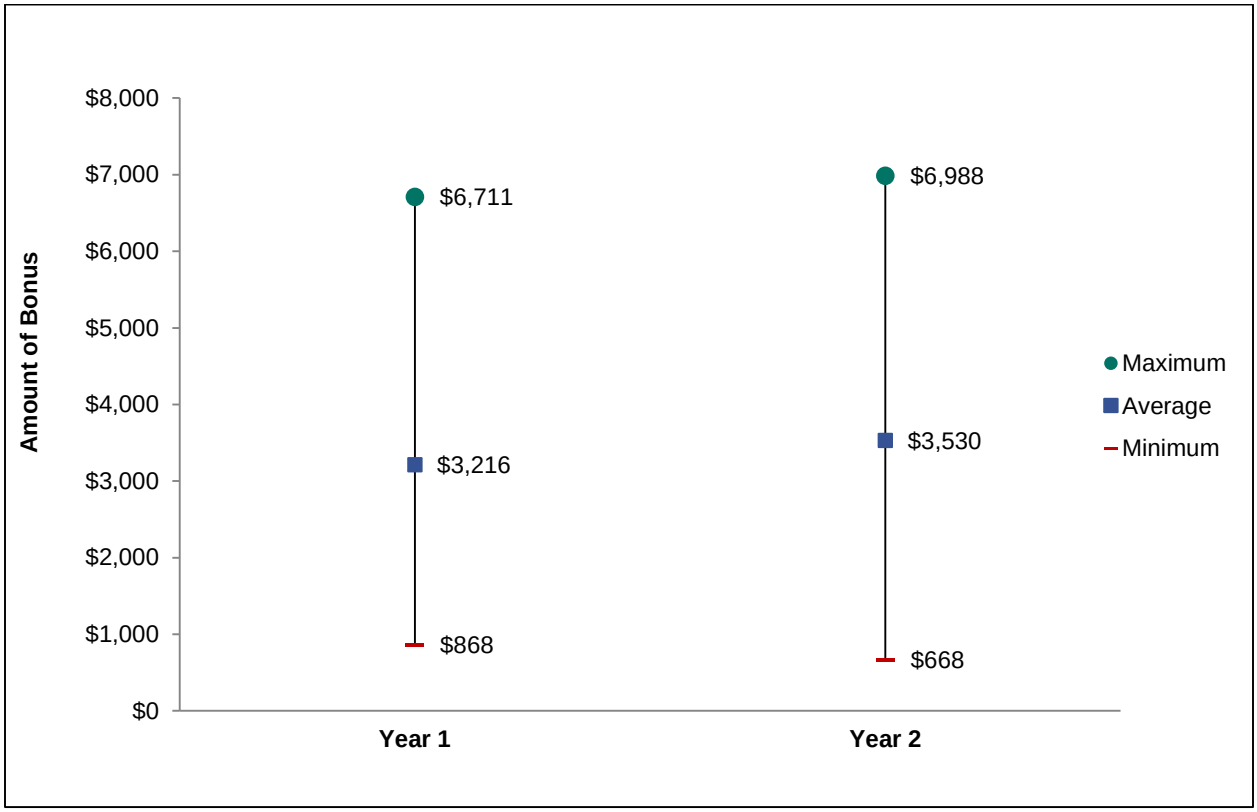
Figure reads: In districts that did not use classroom achievement growth to evaluate teachers in Year 2, 52 percent of the maximum pay-for-performance bonus was based on teachers' classroom observation rating, 8 percent was based on the achievement growth of student subgroups, and 40 percent was based on school achievement growth.

**Table IV.5. Evaluation Districts Meeting TIF Grant Goals for Pay-for-Performance Bonuses for Principals (Percentages)**

| TIF Grant Goal | Year 1 | Year 2 |
|---|---|---|
| Substantial: Average bonus was at least 5 percent of average salary | 30 | 30 |
| Differentiated: Highest bonus was at least three times the average bonus | 10 | 10 |
| Challenging: Less than 50 percent of teachers received a pay-for-performance bonus | 20 | 20 |
| **Number of Districts** | **10** | **10** |

Source: Educator administrative data.

**Figure IV.7. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Principals**




Source: Educator administrative data (N = 65 principals in Year 1 and N = 68 principals in Year 2).

Note: The statistics shown in the figure represent an equal-weighted average of the statistics from the 10 evaluation districts in Cohort 1. When districts were weighted by the number of schools, average bonus amounts were similar to those shown in this figure, but maximum bonus amounts were about $1,200 higher than those shown in this figure (Appendix D, Figure D.10).

Figure reads: In Year 2, on average across the evaluation districts, the minimum pay-for-performance bonus was $668, the average pay-for-performance bonus was $3,530, and the maximum pay-for-performance bonus was $6,988.

**As intended by the study design, the automatic 1 percent bonus provided to teachers and principals in control schools was small and did not vary substantially.** The automatic bonus for educators in control schools ensured that all educators in evaluation schools had the opportunity to monetarily benefit from participating in the study. However, the automatic bonuses were purposefully designed to be small and fairly uniform in order for educators in treatment schools to be eligible for larger and more differentiated bonuses than educators in control schools. The average automatic bonus for teachers in control schools was $402 and $382 in Years 1 and 2, and the maximum automatic bonus was only slightly higher ($635 in Year 1 and $607 in Year 2; Appendix D, Figure D.13). For principals in control schools, the average automatic bonus was $768 and $701 in Years 1 and 2, respectively, with maximum automatic bonuses in these years of $914 and $865. In Years 1 and 2, educators in control schools received automatic bonuses that were, on average, approximately 20 percent of the average amount of the performance bonuses educators in treatment schools received.

### Requirement 3: Additional Pay Opportunities

Consistent with the goal of improving the teaching workforce in high-need schools, the TIF grant required that districts provide additional pay for effective educators to take on extra roles and responsibilities. Examples from the TIF notice included serving as a master or mentor teacher whose

roles typically include mentoring novice teachers, developing professional learning communities, and tutoring students. Using data from district surveys, district interviews, and administrative data, we examined the percentage of evaluation districts that provided additional pay opportunities, the types of roles and responsibilities offered, and the amount of the additional pay.

**All evaluation districts met the TIF grant requirement to offer additional pay opportunities, most commonly in the form of master or mentor/lead teacher opportunities.** All districts reported offering additional pay for teachers to take on extra roles and responsibilities, but none reported offering similar opportunities to principals. Districts most commonly reported offering teachers additional pay for the roles of master and mentor teachers—at least 70 percent of evaluation districts reported offering these roles (Table IV.6). During interviews, officials from districts that offered these roles reported that the number of master teacher positions available within districts ranged from 6 to 61 (depending on the number of study schools) and that the number of mentor teacher positions at each school ranged from 1 to 6. Districts noted that master teachers might lead professional development sessions, and mentor teachers might provide day-to-day coaching or modeling lessons.

We compared the amount of money educators could earn for these additional pay opportunities to the amount they could earn for pay-for-performance bonuses. According to the theory of change (Chapter I), pay-for-performance is expected to encourage teachers to improve their practices in order to receive a bonus. However, if effective teachers could earn as much or more from becoming a master or mentor teacher, then teachers in treatment and control schools might have had similar incentives to improve in order to be qualified for these additional pay opportunities. If so, these additional pay opportunities could have diminished the potential impacts of pay-for-performance.

**Table IV.6. Additional Pay Opportunities for Teachers, as Reported by Districts, Year 2**

| | Percentage of Districts That Offered Additional Pay | Average Maximum Pay in Districts Offering Additional Pay |
|---|---|---|
| Teachers could receive additional pay for taking on extra roles or responsibilities | 100 | NA |
| Roles and Responsibilities | | |
| Mentor teacher | 80 | $3,275 |
| Master or lead teacher | 70 | $9,071 |
| Department chair or head | 20 | — |
| Lead curriculum specialist | 30 | $3,833 |
| Serving on a schoolwide committee or task force | 10 | — |
| Leadership team member | 20 | — |
| Additional Factors | | |
| Teaching in a hard-to-staff school or high-need subject area | 50 | $3,840 |
| Attending professional development activities or enrolling in graduate-level courses | 40 | $413 |
| **Number of Districts—Range[a]** | **10** | **3-8** |

Source: District survey, 2013.

Note: Table reports on activities funded by TIF.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

— is not reported due to small sample size.

NA is not applicable.

Although the maximum pay teachers could earn from taking on additional responsibilities was larger than from pay-for-performance bonuses, they actually earned less, on average, from these additional pay opportunities. In Year 2, the reported maximum additional pay of $9,071 for serving as a master or lead teacher (among evaluation districts offering this type of pay) exceeded the maximum pay-for-performance bonus of $6,894 (among all evaluation districts) (Table IV.6). However, the average actual pay for additional roles and responsibilities in Year 2 was $502, less than 30 percent of the average performance bonus for teachers of $1,760 (Appendix D, Table D.11). This is because only a small fraction of teachers (17 percent) received additional pay for extra work.[42] Additional pay opportunities may also be less attractive than a pay-for-performance bonus if the amount and type of additional work required for the additional pay do not appeal to teachers.

### Requirement 4: Professional Development

The TIF grant required that districts provide professional development linked to the measures of educator effectiveness. This support included professional development to help educators understand the measures being used to evaluate their performance, as well as feedback based on their actual performance ratings to help improve their instructional practices.[43] To describe this required component, we used data from the district survey and interviews with district administrators.

**Most evaluation districts provided the required professional development to teachers.** All evaluation districts offered professional development to help teachers understand the performance measures used for their TIF program. However, only 7 of the 10 districts offered the more targeted professional development based on teachers' own performance ratings (Table IV.7). Although districts reported that a majority of teachers were expected to receive each type of required professional development (Appendix D, Table D.13), half reported that teachers had flexibility in choosing which professional development opportunities they attended (Appendix D, Table D.14).

**Table IV.7. Professional Development Activities for Teachers Planned Under TIF, as Reported by Districts, Year 2 (Percentages)**

| | Evaluation Districts |
|---|---|
| Focus of Professional Development | |
| Understanding performance measures of TIF program | 100 |
| Feedback based on TIF performance ratings | 70 |
| **Number of Districts** | **10** |

Source: District survey, 2013.

## Communication of TIF Program

In addition to determining how to implement the required components of TIF, districts had to effectively communicate information about those components to educators. In this section, we describe evaluation districts' reported communication about their TIF program, including how information was communicated to educators, the timing of the communication, and the content of the information. We focus on two types of information that districts needed to communicate in Year

[42] Average amounts of additional pay for roles and responsibilities did not differ between teachers in treatment and control schools (see Appendix D, Table D.12).

[43] Surveys of district administrators did not ask about professional development for principals.

2: (1) general information about the program, and (2) specific information to individual teachers about the performance bonuses that they earned in Year 1. Data for this section come from the district survey and interviews with district administrators.[44]

**District or grantee staff typically communicated general information about TIF programs to teachers, held in-person meetings to explain the program, and adjusted their communication approaches based on lessons learned.** Deciding who communicates about TIF involves a trade-off. Communication by district or grantee staff might help ensure uniformity and accuracy of information, but communication by school staff (for example, asking principals to explain the program to their teachers) uses staff who might have closer relationships with the teachers.[45] According to reports by district officials during interviews, 7 of 10 districts reported that communication about TIF came from district or grantee staff (Appendix D, Table D.15).

Most districts (90 percent) reported holding in-person meetings to explain the TIF program to teachers (Appendix D, Table D.15). In interviews, district officials also noted using written materials (60 percent), conversations with other stakeholders (40 percent), a district website (30 percent), and media coverage (20 percent).

During interviews, at least six evaluation districts reported that they adjusted an aspect of their communication approach for TIF in 2012–2013 based on lessons learned from the previous school year (Appendix D, Table D.15). For example, district officials said that they added more time for communicating with teachers about TIF (in small groups or during individual meetings) or that they reassigned communication responsibilities.

Districts reported varying in the frequency of their communication activities about TIF. Four districts indicated they communicated with teachers monthly (Appendix D, Table D.15). Of the remaining districts, officials either reported communicating with teachers less than three times during the year or didn't know how often the program was communicated.

**At least half of the districts informed teachers about the actual or expected number, size, and distribution of the pay-for-performance bonuses.** During interviews, 60 percent of the districts reported that they informed teachers about the number, size, and distribution of bonuses awarded for the 2011–2012 school year (Appendix D, Table D.15). Half of the districts reported that they informed teachers about the expected number, size, and distribution of bonuses that would be awarded for the 2012–2013 school year.

**Most districts used letters and email to let teachers know whether they had earned individual performance bonuses and, if so, how much they had earned.** Some methods of communicating about performance bonuses, such as written correspondence, may better ensure uniformity of the message. However, holding individual meetings with teachers to discuss their bonuses may enable teachers to better understand why they received (or did not receive) a bonus. In general, evaluation districts chose uniform written correspondence, rather than individualized in-

[44] The district survey and interview in Year 2 asked district administrators to describe their strategies for communicating Year 1 bonuses. A future report will include information on how districts communicated Year 2 bonuses.

[45] As discussed in Chapter II, 4 of the 10 districts received TIF grants directly from the U.S. Department of Education. The remaining districts were part of multidistrict grants administered by another grantee organization (such as a state education agency, university, association of charter schools, or nonprofit organization), and grantee or district staff could have helped ensure uniformity of the information communicated to educators.

person meetings, to inform teachers of their individual bonuses. Most of the evaluation districts (80 percent) reported informing teachers about their individual performance bonus by sending a letter or email to the teacher. Thirty percent reported holding individual meetings with teachers to discuss the bonus amount they received (Table IV.8).

When informing teachers of their individual bonuses, most districts (6 of 10) reported that they also reminded educators of the criteria for earning bonuses. For example, district officials described giving educators information on the maximum bonus amount available for a performance rating or providing an algorithm for how individual bonuses were determined.

**Districts used group presentations to inform teachers about school-level bonuses.** Ninety percent of the districts reported using group presentations to inform teachers about school- or district-level bonus amounts. During interviews, district officials from at least four districts explained that they used this approach to ensure thorough and consistent communication.

Another decision that districts made was whether to inform teachers who failed to earn a bonus that they did not earn one. Doing so could have helped ensure that those nonrecipients were aware of the missed opportunity to earn a bonus and motivate them to improve their teaching practices. During interviews, administrators from four districts indicated that nonrecipients were informed that they did not earn a bonus, and administrators from two districts reported that no such information was provided to nonrecipients. Administrators in the remaining four districts did not indicate what information was provided to nonrecipients because they expected all eligible teachers to receive a bonus.

**Table IV.8. Communication Methods Used to Inform Teachers and Other Stakeholders About Pay-for-Performance Bonuses Based on the First Year of TIF Implementation (Percentages)**

| | Percentage of Evaluation Districts |
|---|---|
| Approaches for informing teachers about their individual bonus amounts | |
| Letter or email to each teacher with individual bonus amount | 80 |
| Individual meeting with each teacher to discuss bonus amount | 30 |
| Approaches for informing groups of staff about school or district bonus amounts | |
| Letter or email with school- or district-level information about bonus amount | 50 |
| Group presentation to teachers with school- or district-level information about bonus amount | 90 |
| Approaches for publicly reporting bonus amounts | |
| Presentation to school board | 40 |
| Press release or press coverage about bonus amount for individual teachers | 0 |
| Press release or press coverage with school- or district-level information about bonus amount | 10 |
| District website | 20 |
| **Number of Districts** | **10** |

Source: District survey, 2013.

**Most districts did not notify teachers of the bonuses they earned for Year 1 before the start of the next school year.** For information about bonuses to affect teachers' behavior, teachers must receive the information when there is still enough time to affect their school choice (for example, requesting a transfer to a school that offers or does not offer a bonus) or their teaching practices (for example, enrolling in professional development to learn how to perform better on the performance measures used to award bonuses). For the 9 of 10 districts that awarded any bonuses based on

teachers' performance in 2011–2012, there were differences in the timing with which districts notified teachers of their bonuses and paid out those bonuses. Only three districts reported notifying and paying any teachers before the start of the 2012–2013 school year. The remaining six districts reported notifying and paying teachers between October and December 2012.

Even if districts did not notify and award bonuses before the start of the next school year, teachers may have known their scores on observation measures earlier. For at least three districts, notification and payment of bonuses occurred earlier for bonuses based on observations of classroom or school practices than for bonuses based on achievement growth measures. For example, during interviews, district officials reported that educators learned whether they got a performance bonus based on observations at the end of the school year (usually when their summative scores were discussed), but they learned about bonuses based on achievement growth between one and six months later.

## Teacher and Principal Perspectives Regarding TIF Implementation

Teachers' and principals' understanding of the TIF program is important because it reflects how well the program's incentives were communicated and, in turn, can determine how the program may influence educators' behaviors and, ultimately, student achievement (as described by the theory of change discussed in Chapter I). Moreover, educators' reports about program features can identify ways in which their understanding of the TIF program did or did not align with what grantees intended or what district officials reported, highlighting possible challenges in the implementation process.

This section examines educators' reported understanding of and experiences with TIF performance measures, pay-for-performance bonuses, additional pay opportunities, and professional development, drawing primarily on teachers' and principals' survey responses. Although pay-for-performance was the only component that was supposed to differ between treatment and control schools, educators' understanding of all four required components could have differed between treatment and control schools (if, for example, information was communicated differently to the two groups of educators or they paid different amounts of attention to this information). Therefore, we describe the perspectives of treatment and control educators separately. We also examine educators' evolving understanding of their TIF program, because that understanding might change as districts refine communication strategies and information becomes more widely disseminated. Because we administered Year 1 surveys before educators had received any performance bonuses and administered Year 2 surveys after Year 1 bonuses had been awarded, changes in understanding might also result from educators' having received the bonuses or heard about them.

### Educators' Understanding of Performance Measures

For the program to change educators' behavior and, ultimately, student outcomes, educators need to understand how they are being evaluated and how they can change their practices to improve their performance.

**Teachers' understanding of performance measures improved from Year 1 to 2.** For example, about 85 percent of teachers (87 and 84 percent of treatment and control teachers, respectively) reported being evaluated on at least two classroom observations in Year 2 compared to about 75 percent of teachers in Year 1 (Table IV.9).

**In Year 2, educators in treatment schools were more likely than educators in control schools to report being evaluated on student achievement growth.** Consistent with the study

design that only the offer of pay-for-performance bonuses should differ between treatment and control schools, similar percentages of treatment and control teachers reported being evaluated on student achievement growth in Year 1 (about 70 percent). However, by Year 2, a higher percentage of treatment teachers (78 percent) reported being evaluated on student achievement growth, which was also significantly greater than the percentage of control teachers (72 percent) who reported being evaluated on student achievement growth (Table IV.9). Treatment principals also were more likely than control principals in Year 2 to report being evaluated on student achievement growth (91 versus 67 percent). However, this was due to a significant decrease in the percentage of control principals reporting they were being evaluated on student achievement growth in Year 2 (67 percent) compared to Year 1 (92 percent) (Table IV.10).[46]

**Table IV.9. Teachers' Reports of the Measures Used to Evaluate Teachers (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Student Achievement Measures | | | | | | |
| Student achievement level (e.g., percent proficient) | 56 | 61 | -5 | 69+ | 67 | 1 |
| Student achievement growth | 71 | 70 | 2 | 78+ | 72 | 6* |
| By school | 62 | 63 | -1 | 73+ | 68 | 5* |
| By student subgroups[a] | 55 | 56 | -1 | 66+ | 60 | 6* |
| By teacher's classroom | 60 | 62 | -2 | 57 | 58 | -1 |
| Classroom Observation Measure | | | | | | |
| At least two classroom observations by trained observers | 74 | 77 | -3 | 87+ | 84+ | 3 |
| **Number of Teachers—Range[b]** | **382-384** | **394-398** | **513-517** | **432-437** | **432-434** | **443-448** |

Source: Teacher survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Examples of student subgroups include grouping students by grade, team, or subject area.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference between treatment and control group is statistically significant at the 0.05 level, two-tailed test.
+Difference between Year 1 and Year 2 is significant at the 0.05 level, two-tailed test.

[46] When we restricted the sample to principals who responded to the survey in both years, the results were nearly identical. Therefore, the drop in the percentage of control principals reporting that they were evaluated on student achievement growth was not due to a change in which principals responded to the survey.

**Table IV.10. Principals' Reports of the Measures Used to Evaluate Principals (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Student Achievement Measure | | | | | | |
| Student achievement level (e.g., percent proficient) | 90 | 93 | -5 | 85 | 69+ | 16* |
| Student achievement growth | 88 | 92 | -3 | 91 | 67+ | 25* |
| By school | 89 | 90 | -1 | 90 | 65+ | 25* |
| By student subgroups[a] | 83 | 90 | -6 | 83 | 64+ | 19* |
| Observation Measure | | | | | | |
| At least two observations by trained observer | — | — | — | 60 | 73 | -13 |
| **Number of Principals—Range[b]** | **59-63** | **58-60** | | **63-64** | **56-58** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Examples of student subgroups include grouping students by grade, team, or subject area.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

— is not available.

*Difference between treatment and control group is statistically significant at the 0.05 level, two-tailed test.
+Difference between Year 1 and Year 2 within treatment status is statistically significant at the 0.05 level, two-tailed test.

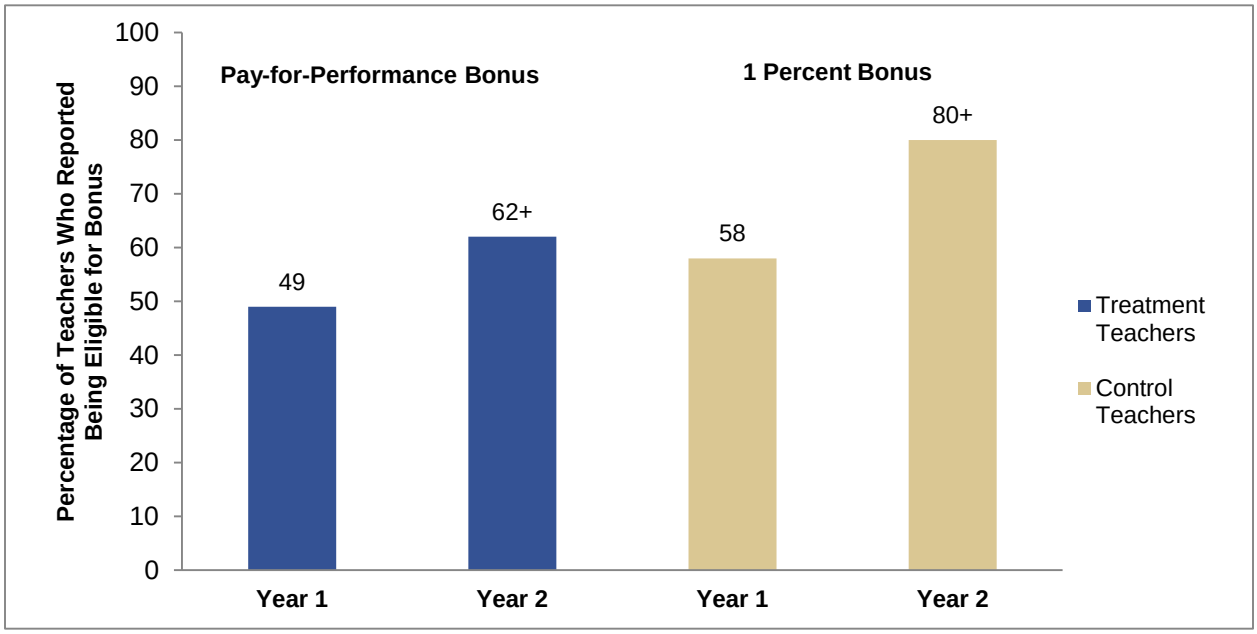
## Educators' Understanding of Their Eligibility for Pay-for-Performance Bonuses

The prospect of earning a performance bonus could motivate educators to improve their practice. To do so, however, they need to have a correct understanding of their eligibility for bonuses. Based on the study design, we would expect that all teachers in treatment schools would report being eligible for a pay-for-performance bonus, while teachers in control schools would report only being eligible for an automatic 1 percent bonus.

**Understanding of bonus eligibility improved among both teachers and principals, but many teachers continued to misreport their eligibility.** In Year 2, 62 percent of teachers in treatment schools reported they were eligible for a pay-for-performance bonus, and 80 percent of teachers in control schools reported they were eligible for an automatic 1 percent bonus. Between Years 1 and 2, there was a 13 percentage point increase in the percentage of treatment teachers who reported they were eligible for a pay-for-performance bonus and a 22 percentage point increase in the percentage of control teachers who reported they were eligible for an automatic 1 percent bonus (Figure IV.8). Nearly all principals in treatment schools (90 percent) said they were eligible for a pay-for-performance bonus, and 85 percent of principals in control schools said they were eligible for an automatic 1 percent bonus. There was a 35 percentage point increase in the percentage of principals in treatment schools who reported they were eligible for a pay-for-performance bonus and a 19

percentage point increase in the percentage of principals in control schools who reported being eligible for an automatic 1 percent bonus (Figure IV.9).[47,48]

**Figure IV.8. Teachers' Bonus Eligibility, as Reported by Teachers**




Source: Teacher surveys, 2012 and 2013.

Notes: A total of 377 treatment teachers in Year 1 and 444 in Year 2 responded to the question about eligibility for a pay-for-performance bonus. A total of 381 control teachers in Year 1 and 445 in Year 2 responded to the question about eligibility for an automatic 1 percent bonus.

Figure reads: Among teachers in treatment schools, 49 and 62 percent reported being eligible for a pay-for-performance bonus in Year 1 and Year 2, respectively.

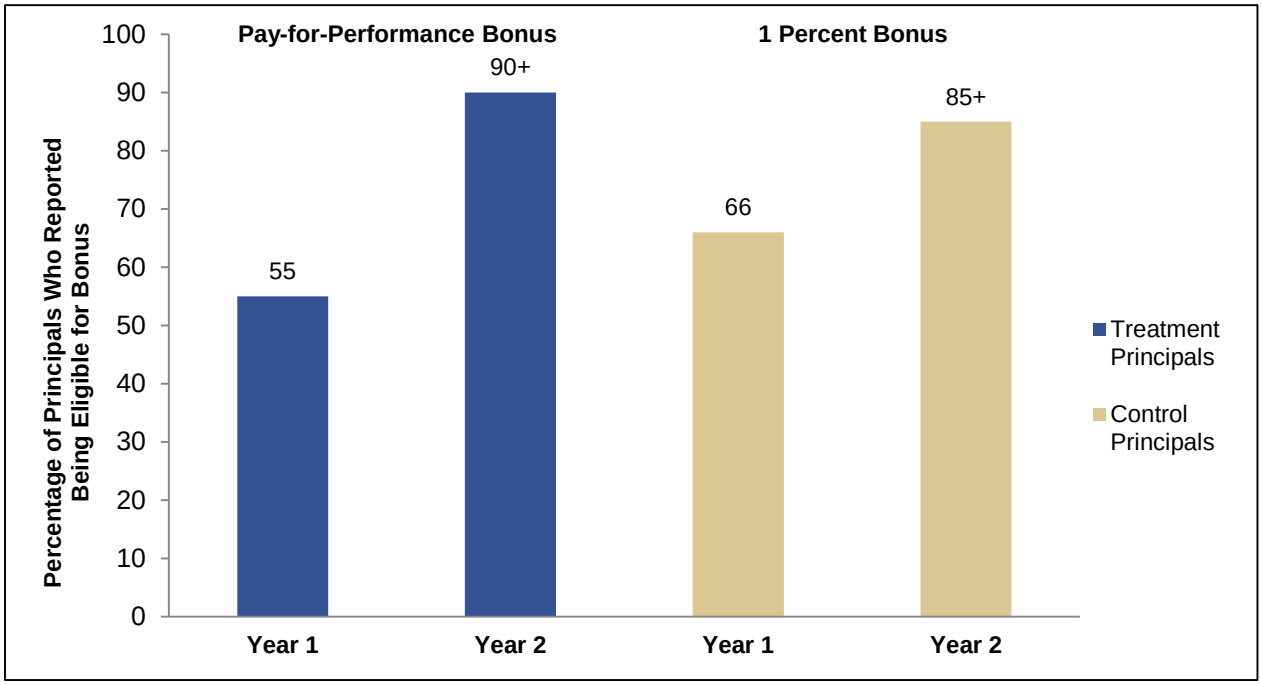
+ Difference between Year 1 and Year 2 within treatment status is statistically significant at the 0.05 level, two-tailed test.

Because understanding about eligibility for a bonus is critical for changing behavior, we explored how teacher understanding varied across districts, across schools within the same district, and within the same school. If teacher understanding did not vary within a district, we might hypothesize that districtwide factors, such as whether bonuses were included in teachers' regular paychecks or in separate bonus paychecks, were important in determining teachers' understanding. If teacher understanding varied within a district, but not within a school, we might conclude that school factors, such as whether the principal correctly understood and conveyed teachers' eligibility, influenced teachers' understanding. If teacher understanding varied within a school, variation in teachers' understanding may be explained by differences in teachers' characteristics, such as whether the teacher had ever received a bonus or whether the teacher attended TIF-related professional development sessions.

[47] When Year 1 analyses were based on Cohorts 1 and 2, similar, but somewhat smaller, percentages of teachers and principals reported being eligible for the correct type of bonus (Appendix D, Figures D.14 and D.15).

[48] Some educators thought they were eligible for the wrong bonus. For example, in Year 2, 18 percent of control teachers and 15 percent of control principals thought they were eligible for pay-for-performance bonuses (Appendix D, Table D.16).

**Figure IV.9. Principals' Bonus Eligibility, as Reported by Principals**




Source: Principal surveys, 2012 and 2013.

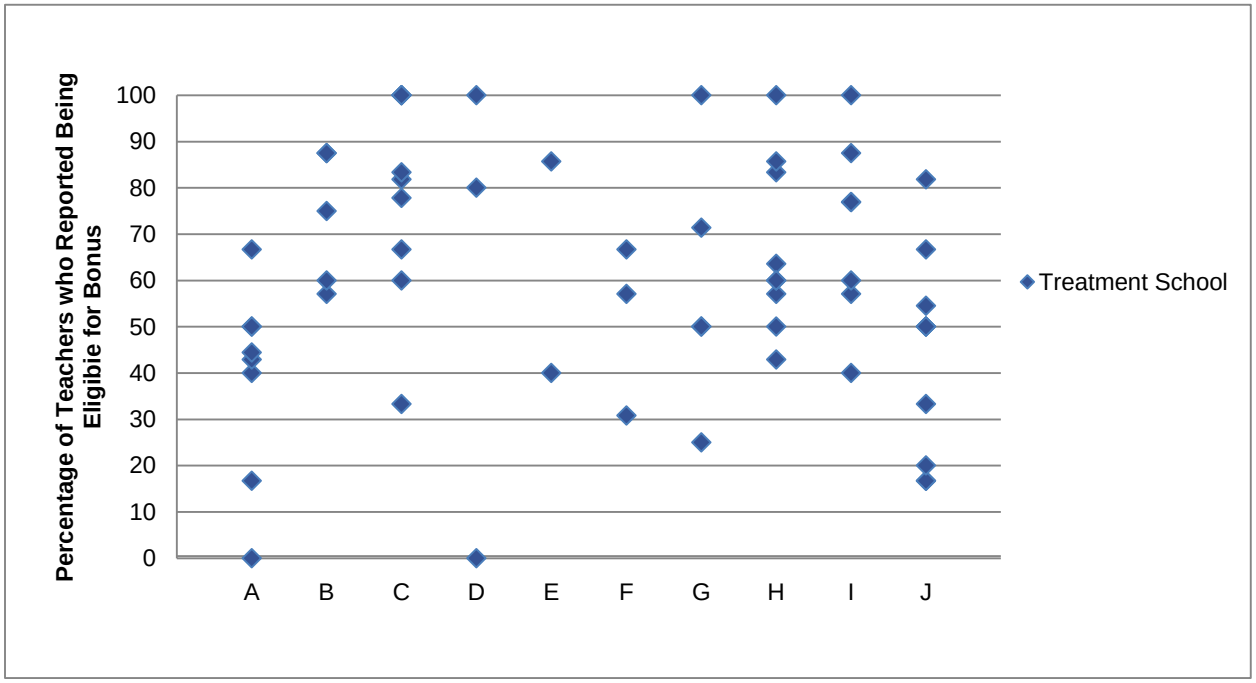
Notes: A total of 64 treatment principals in Year 1 and 63 in Year 2 responded to the question about eligibility for a pay-for-performance bonus. A total of 64 control principals in Year 1 and 61 in Year 2 responded to the question about eligibility for an automatic 1 percent bonus.

Figure reads: Among principals in treatment schools, 55 and 90 percent reported being eligible for a pay-for-performance bonus in Year 1 and Year 2, respectively.

+ Difference between Year 1 and Year 2 within treatment status is statistically significant at the 0.05 level, two-tailed test.

**Most of the differences in teachers' understanding occurred among teachers in the same school.** Figure IV.10 displays the variation in treatment teachers' understanding of eligibility for pay-for-performance bonuses for each evaluation district. Each diamond on the figure represents a treatment school and shows the percentage of teachers in that school reporting they were eligible for a performance bonus. A diamond at the top of the figure (100) indicates that all the teachers in that school correctly reported being eligible for a pay-for-performance bonus. As the figure shows, teacher understanding varied within districts and within schools. In fact, in many treatment schools, about half of the teachers reported being eligible for pay-for-performance bonuses, and half reported they were not eligible. Statistically, we found that more than 85 percent of the variation in treatment teachers' understanding of their eligibility for a pay-for-performance bonus occurred among teachers in the same school (Appendix D, Table D.17).

**Figure IV.10. Treatment Teachers' Reported Pay-for-Performance Bonus Eligibility by School and by District, Year 2 (Percentages)**




Source: Teacher survey, 2013 (N = 435 teachers in 61 treatment schools).

Figure reads: In Year 2, of the two schools in District E that offered teachers pay-for-performance bonuses, 40 percent of teachers in one school reported being eligible for a bonus and about 85 percent of teachers in the other school reported being eligible for a bonus.

We examined a variety of district, program, teacher, and school characteristics to determine whether differences in these factors could help explain differences in treatment teachers' understanding of their eligibility for a performance bonus. The district and program characteristics we examined included whether the district (1) had another additional compensation program, (2) adjusted its communication approach based on lessons learned from the prior year, (3) used district or school staff to communicate the TIF program to teachers, (4) assessed teachers' understanding using focus groups or surveys, (5) implemented TAP[49], (6) expected at least 75 percent of teachers to attend TIF-required professional development, (7) paid pay-for-performance bonuses through teachers' regular paycheck (rather than a separate check), (8) communicated the number, size, and distribution of actual bonuses awarded in Year 1, and (9) communicated expectations about the number, size, and distribution of bonuses to be awarded in Year 2. Teacher characteristics we examined included whether the teacher (1) taught a tested grade/subject, (2) received a performance bonus for Year 1, (3) participated in TIF-related professional development, and (4) was or had a mentor teacher. We also examined one school factor—principals' understanding of teachers' eligibility.

**Teachers in districts that adjusted their communication had a better understanding of their eligibility for performance bonuses.** A higher percentage of treatment teachers (73 percent) in districts that adjusted their communication based on lessons learned from Year 1 reported they were eligible for pay-for-performance bonuses than treatment teachers (57 percent) working in districts that did not adjust their communication strategy (Appendix D, Table D.18). None of the

[49] The Teacher Advancement Program (TAP) is a comprehensive teacher pay reform model.

other characteristics we examined could account for the variation in teachers' understanding (Appendix D, Tables D.18 and D.19).

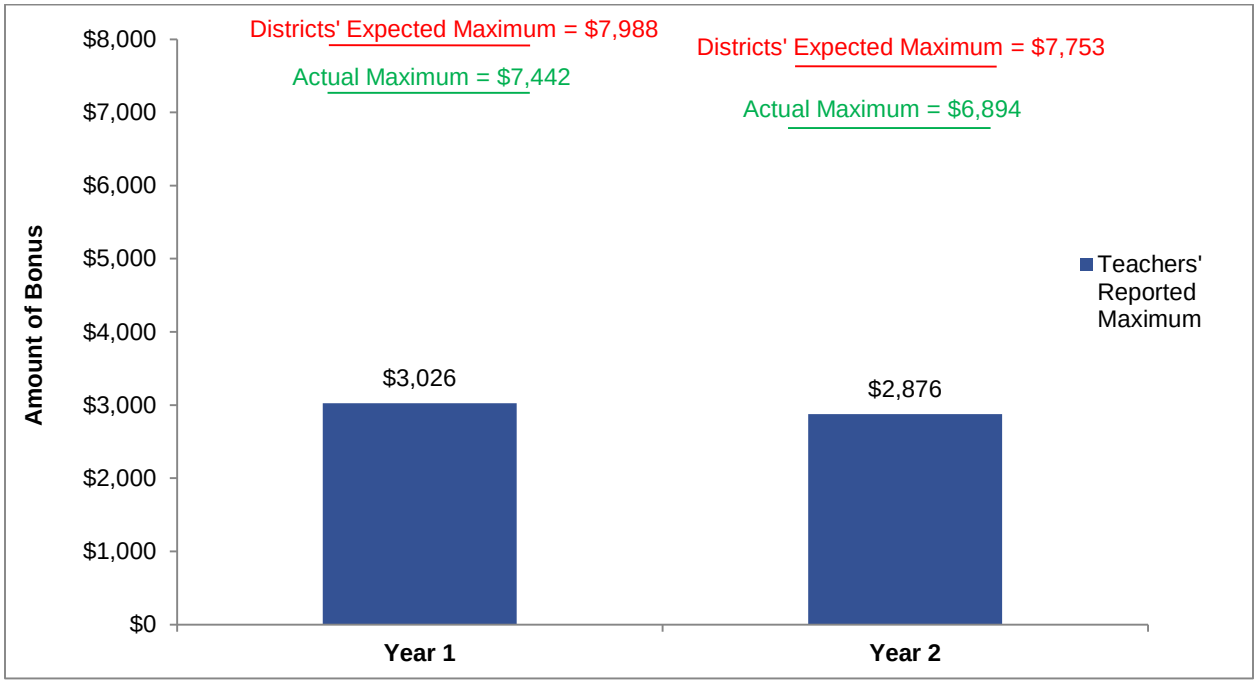
**Educators' Understanding of the Potential Amounts of Pay-for-Performance Bonuses**

For performance bonuses to provide an incentive for teachers to change their behaviors, teachers not only need to understand they are eligible for a bonus, but they must also believe the potential amount of the bonus is enough to change their teaching practices or effort. Figure IV.11 shows, on average across districts, the maximum performance bonus that teachers believed was available, the maximum performance bonus that districts reported teachers could earn, and the actual maximum performance bonus that was awarded to teachers. Teachers' expectations in Year 1 would have been primarily shaped by how well districts communicated the design of the pay-for-performance component to their teachers. By Year 2, however, teachers' expectations could also have been influenced by the actual bonuses awarded after Year 1.

**Teachers underestimated the maximum amount of performance bonuses throughout the first two years of the TIF program.** In fact, treatment teachers reported no better understanding of the maximum performance bonus in Year 2 than in Year 1. In Year 2, teachers in treatment schools, on average, reported that the maximum pay-for-performance bonus they could receive was $2,876. This amount was about two-fifths the size of the actual maximum bonus of $6,894 awarded by districts (Figure IV.11). In comparison, in Year 1, teachers in treatment schools reported that the maximum pay-for-performance bonus they could receive was $3,026, which was about two-fifths the size of the actual maximum bonus ($7,442). The average maximum pay-for-performance reported by teachers includes a maximum bonus of $0 for teachers who did not believe they were eligible for a performance bonus. Therefore, the maximum bonus reported by teachers, on average, may have been lower than the maximum reported by the district because of teachers' misunderstanding of their eligibility. However, even the teachers who believed they were eligible for a performance bonus underestimated the potential amount, reporting that, on average, the maximum performance bonus they could receive was about $3,600 in both years (not shown).

**Principals also continued to underestimate the potential amount of performance bonuses they could receive, but their expectations were better aligned with actual bonus payouts than were teachers' expectations.** In Year 2, principals in treatment schools, on average, reported that the maximum pay-for-performance bonus they could receive was $6,097. This amount was 87 percent of the actual maximum bonus ($6,988) awarded to principals (Figure IV.12).

**Figure IV.11. Actual and Reported Maximum Pay-for-Performance Bonus for Teachers in Treatment Schools**



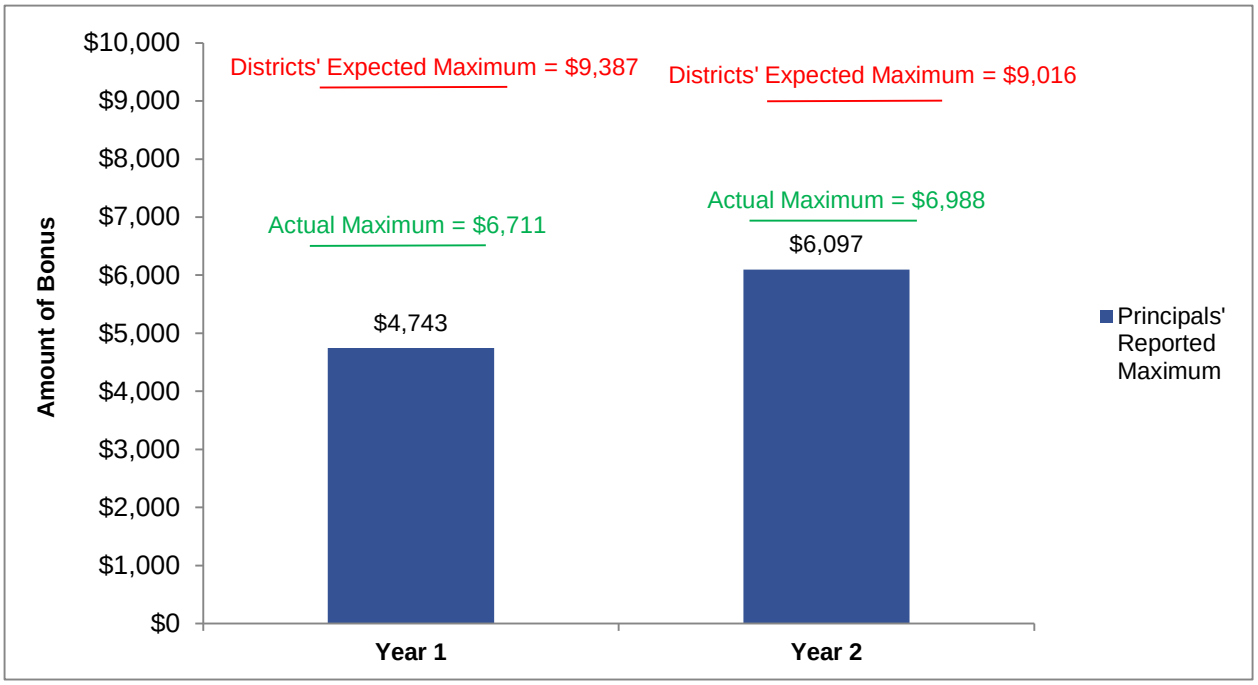

Source: Teacher survey (2012 and 2013), district interviews (2012 and 2013), and administrative data.

Notes: Teachers' reports are based on data for teachers in tested grades and subjects, with each school receiving an equal weight. Districts' reports and payouts are based on data for all teachers, with each district receiving an equal weight. Appendix D, Figure D.16 shows that our results are similar if all reports are based on giving schools equal weight.

A total of 196 treatment teachers and 214 control teachers in tested grades and subjects responded to this survey question in Year 1. A total of 218 treatment teachers and 246 control teachers in tested grades and subjects responded to this survey question in Year 2. The maximum bonus amount was set to zero for all respondents who indicated they were ineligible for a bonus. For teachers who reported being eligible for the bonus but left the amount missing, bonus amounts were imputed through multiple imputation methods. This led to 27 additional responses for treatment teachers and 7 for control teachers in Year 1 and to 14 additional responses for treatment teachers and 6 for control teachers in Year 2. See Appendix B for additional discussion on the imputation methods. Appendix D, Table D.20 shows that our results are similar if we do not impute the missing bonus amounts.

Figure reads: In Year 2, on average the evaluation districts expected that the maximum pay-for-performance bonus that a teacher could earn was $7,753, the actual maximum pay-for-performance bonus awarded was $6,984, and the maximum pay-for-performance bonus teachers reported they could earn was $2,876.

**Figure IV.12. Actual and Reported Maximum Pay-for-Performance Bonus for Principals in Treatment Schools**




Source: Principal survey (2012 and 2013), district interviews (2012 and 2013), and administrative data.

Note: Principals' reported values were calculated giving each school an equal weight. Districts' reports and payouts were calculated giving each district an equal weight. When districts were weighted by the number of schools, actual maximum performance bonus amounts for principals were higher ($7,884 in Year 1 and $8,191 in Year 2), implying a somewhat wider gap between principals' reported maximum bonus amounts and the actual amounts (Appendix D, Figure D.17).

A total of 56 treatment principals and 60 control principals responded to this survey question in Year 1. A total of 61 treatment principals and 61 control principals responded to this survey question in Year 2. The maximum bonus amount was set to zero for all respondents who indicated they were ineligible for a bonus. For educators who reported being eligible for the bonus but left the amount missing, bonus amounts were imputed through multiple imputation methods. This led to 8 additional responses for treatment principals and 3 for principals in Year 1 and to 2 additional responses for treatment principals and 0 for control principals in Year 2. See Appendix B for additional discussion on the imputation methods. Appendix D, Table D.20 shows that our results are similar if we do not impute the missing bonus amounts.

Figure reads: In Year 2, on average the evaluation districts expected that the maximum pay-for-performance bonus that a principal could earn was $9,016, the actual maximum pay-for-performance bonus awarded was $6,988, and the maximum pay-for-performance bonus principals reported they could earn was $6,097.

## Educators' Understanding of and Experiences with Other Required Components

Educators also reported their understanding of and experiences with the remaining two required components: (1) additional pay opportunities, and (2) professional development to help them understand and improve their ratings on TIF performance measures. Educators' understanding of additional pay opportunities can shed light on how visible these opportunities were in the study schools. Educators' reported participation in TIF-related professional development can suggest the extent to which districts allocated resources and attention to this component. It may also shed light on whether educators received enough guidance to know how to improve their performance. As with implementing the performance measures used in TIF, evaluation districts were expected to implement these required components identically in treatment and control schools.

**Teachers' awareness of additional pay opportunities increased from Year 1 to 2.** In Year 2, nearly 90 percent of teachers reported that opportunities for earning extra pay for additional roles and responsibilities were available at their school. Compared to Year 1, these percentages were significantly higher for teachers in both treatment and control schools (about 56 percent in Year 1, compared to nearly 90 percent in Year 2 for both treatment and control teachers; Table IV.11). Similar percentages of treatment and control teachers reported availability of additional pay opportunities.

**Table IV.11. Eligibility for Additional Pay Opportunities, as Reported by Teachers and Principals (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| **Teachers** | | | | | | |
| Teachers could receive additional pay for taking on extra roles or responsibilities | 57 | 56 | 1 | 89+ | 88+ | 1 |
| Roles or Responsibilities | | | | | | |
| Mentor teacher | 44 | 40 | 4 | 72+ | 74+ | -2 |
| Master or lead teacher | 40 | 39 | 0 | 54+ | 57+ | -3 |
| Department chair or head | 18 | 20 | -1 | 22 | 29+ | -8* |
| Lead curriculum specialist | 26 | 25 | 1 | 35+ | 38+ | -3 |
| Schoolwide committee or task force member | 11 | 11 | 0 | 18+ | 21+ | -3 |
| Leadership team member | 35 | 29 | 6 | 23+ | 27 | -4 |
| Additional Factors | | | | | | |
| Teach in a hard-to-staff or high-need school | 25 | 24 | 1 | 30+ | 31+ | -1 |
| Attend professional development activities or enroll in graduate level courses | 30 | 28 | 2 | 25 | 24 | 0 |
| **Number of Teachers—Range[a]** | **246-385** | **234-393** | | **450-454** | **450-456** | |
| **Principals** | | | | | | |
| Principals could receive additional pay for taking on extra roles or responsibilities | 2 | 14 | -13* | 20+ | 16 | 4 |
| **Number of Principals** | **64** | **63** | | **64** | **61** | |

Source: Teacher and principal surveys, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference between treatment and control group within year is statistically significant at the 0.05 level, two-tailed test.
+Difference between Year 1 and Year 2 within treatment status is statistically significant at the 0.05 level, two-tailed test.

**Principals were less likely than teachers to report being offered additional pay opportunities.** About 20 percent of principals reported that opportunities for earning extra pay for additional roles and responsibilities were available for principals in Year 2 (Table IV.11). Although none of the districts reported offering extra pay for principals to accept additional responsibilities, 10 percent of the districts reported offering principals extra pay for working in a hard-to-staff school, attending professional development, or enrolling in graduate courses (not shown). Some principals may have interpreted their eligibility for earning extra pay for these other factors as extra pay for additional roles or responsibilities.

**More than half of teachers reported they received the professional development required under the TIF grant but indicated they received only a few hours of it.** In Year 2, approximately two-thirds of teachers reported that they received or expected to receive professional development focused on understanding performance measures used in TIF, and about 55 percent reported receiving or expecting to receive feedback based on their performance ratings (Appendix D, Table D.21). Of those who expected to receive any professional development on these two topics, the expected amount of time on each topic was about four hours (Appendix D, Table D.22).

## Summary

According to the theory of change presented in Chapter I, some key steps needed to occur in the implementation of TIF for pay-for-performance to be able to improve educator effectiveness and student achievement. This chapter examined whether and how each of these steps materialized in the evaluation districts' implementation of TIF. Describing the implementation of the TIF grant in evaluation districts is useful context for interpreting findings presented later in this report on the program's impacts on educator and student outcomes.

The findings from this chapter indicate that evaluation districts made progress in creating the conditions needed for pay-for-performance to improve educator effectiveness and student achievement. By the second year of implementation, the evaluation districts had implemented most of the TIF required components, and implementation and educator understanding improved between Years 1 and 2. For example, by Year 2 all of the evaluation districts measured teacher and principal effectiveness as required, offered teachers and principals performance-based bonuses, and offered educators additional pay opportunities. Also, more teachers understood how they were being evaluated, and a higher percentage of teachers and principals correctly reported being eligible for a pay-for-performance bonus in Year 2.

However, the findings in this chapter also suggest possible factors that may have dampened the potential for pay-for-performance to improve educator effectiveness and student achievement in the first two years of implementation. For example, many teachers in treatment schools continued to believe they were ineligible for a performance bonus or underestimated how much they could earn from these bonuses. It is unlikely that educators will seek to change their practices if they do not believe they are eligible for a performance bonus or believe they can only earn a relatively small bonus. Even if educators had perfect understanding of their eligibility and the amount they could earn, we also found that most educators received a bonus and the average bonuses were not large. Therefore, the actual structure of the bonuses may not have provided educators with an incentive to change their behavior.

If educators were motivated to change their practices, they still may have found it difficult to determine what practices needed adjustment. We found that most teachers received a high score based on observations, but most received a low score based on student achievement growth. Although different performance measures may be designed to evaluate different aspects of performance, teachers may have had trouble deciding whether and how to adjust their teaching practices if they received conflicting information about their performance. In addition, teachers who only received an achievement growth rating based on school achievement growth (rather than the achievement growth based only on the performance of students they teach) may not necessarily understand how their performance contributed to their student achievement growth rating and how or if they could affect that rating.

Finally, the purpose of the professional development requirement of the TIF program was to ensure teachers understood how they were being evaluated and how to change their practices to improve on the measures. We found that teachers received relatively little professional development to help them understand how to change their practices based on their measured performance. For example, 3 of the 10 evaluation districts indicated that they did not provide this type of professional development to teachers. Among teachers who expected to receive any professional development on these two topics, the expected amount of time on each topic was about four hours over the school year.

# V. IMPACTS OF PAY-FOR-PERFORMANCE ON EDUCATORS' ATTITUDES AND BEHAVIORS

The ways in which pay-for-performance programs affect educators' attitudes (such as job satisfaction) and behaviors (such as allocation of time during the day) can shape how pay-for-performance affects student outcomes. As the theory of change in Chapter I shows, pay-for-performance bonuses may improve student achievement by making educators more productive and by attracting and retaining more effective teachers. However, if the presence of pay-for-performance discourages useful collaboration, lowers morale, or makes a school less appealing to effective teachers, it could have a negative effect on the work environment and on student achievement.

In this chapter, we use data from teacher and principal surveys to estimate the impacts of pay-for-performance on educators' self-reported attitudes and behaviors after one and two years of TIF implementation. Educators in treatment schools were eligible for pay-for-performance bonuses, and educators in control schools were not. Because both treatment and control schools offered all the other required components of the TIF program, any differences in responses between educators in treatment schools and control schools can be attributed to the impacts of pay-for-performance.[50]

**Key Findings on the Impacts of Pay-for-Performance on Educators' Attitudes and Behaviors**

- **Most teachers and principals reported being satisfied with their professional opportunities, some factors associated with how they were evaluated, and their school environment.**
- **Teachers in treatment schools were less satisfied than teachers in control schools with some factors associated with how they were evaluated and school environment, but were more satisfied with their opportunities to earn extra pay.**
- **Principals in treatment schools were less satisfied than principals in control schools with the measures used to evaluate their performance.**
- **Most teachers and principals had positive attitudes toward the TIF program, but teachers in treatment schools were less likely than teachers in control schools to be positive about TIF.**
- **Experienced teachers responded least favorably to pay-for-performance.**

The chapter is based on 10 evaluation districts that completed two years of TIF implementation during the period covered by this report. We refer to the first and second years of implementation, 2011–2012 and 2012– 2013, as Years 1 and 2.[51] Although attitudes and behaviors in Year 2 are a key focus of this chapter, we also examined how these outcomes evolved between Years 1 and 2. Year 1 surveys were administered before educators had received any performance bonuses, whereas Year 2

[50] As discussed in Chapter IV, some educators in the study schools misunderstood their eligibility for pay-for-performance or the potential amounts they could earn. The impacts reported in this chapter reflect the impact of pay-for-performance given educators' actual beliefs. This study was not designed to assess the impacts of pay-for-performance bonuses if all educators correctly understood their eligibility or the amount they could earn in a bonus.

[51] As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts examined in this chapter, whose schools were randomly assigned in spring and summer 2011, were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort. In Appendix E, Tables E.1 through E.4, we present impacts on educators' satisfaction and attitudes from Year 1 for Cohorts 1 and 2 together—that is, findings from 2011–2012 for Cohort 1 and from 2012–2013 for Cohort 2.

surveys were administered after Year 1 bonuses had been awarded. Therefore, these data provide an opportunity to examine whether educators' initial impressions of performance-based compensation changed as bonuses were awarded and educators gained more experience with the program components.

## Impact of Pay-for-Performance on Educators' Attitudes

In this section, we present estimates of the impact of pay-for-performance on educators' satisfaction and attitudes toward their jobs and the TIF program.

### Satisfaction with Job and Factors Associated with Evaluation System

**Most teachers in both treatment and control schools were satisfied with their professional opportunities, factors associated with how they were evaluated, and school environment.** In Year 2, at least 80 percent of teachers reported being somewhat or very satisfied with their opportunities to enhance their skills, their quality of interaction with colleagues, and colleagues' efforts (Table V.1). Teachers reported being least satisfied with opportunities to earn extra pay (62 percent of treatment teachers and 54 percent of control teachers) and school morale (58 percent of treatment teachers and 59 percent of control teachers).

**Table V.1. Teachers' Satisfaction with Professional Opportunities, Evaluation System, and School Environment (Percentages Who Are "Somewhat" or "Very" Satisfied)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Satisfaction Dimension | Treatment | Control | Impact | Treatment | Control | Impact |
| Opportunities for Pay and Development | | | | | | |
| Opportunities for professional advancement | 67 | 75 | -9* | 72 | 74 | -3 |
| Opportunities to enhance skills | 76 | 78 | -2 | 80 | 81 | -1 |
| Opportunities to earn extra pay | 62 | 57 | 6 | 62 | 54 | 9* |
| Factors Associated with Evaluation System | | | | | | |
| Use of student achievement scores to assess performance | 66 | 67 | -1 | 60 | 69 | -9* |
| Feedback on my performance | — | — | — | 75 | 80 | -5* |
| School Environment | | | | | | |
| Recognition of accomplishments | 54 | 61 | -7* | 60 | 66 | -6* |
| Quality of interaction with colleagues | 75 | 81 | -6* | 82+ | 82 | 0 |
| Colleagues' efforts | 83 | 85 | -1 | 84 | 83 | 0 |
| School morale | 50 | 55 | -5 | 58 | 59 | -1 |
| Job Satisfaction | | | | | | |
| Overall job satisfaction | 68 | 73 | -5 | 73 | 74 | -1 |
| **Number of Teachers—Range[a]** | **387-391** | **392-399** | | **444-448** | **446-449** | |

Source: Teacher survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

— is not available.

*Impact is statistically significant at the .05 level, two-tailed test.
+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

**Teachers in treatment schools were less satisfied than teachers in control schools with some factors associated with how they were evaluated and school environment, but were more satisfied with their opportunities to earn extra pay.** In Year 2, a lower percentage of teachers in treatment schools than control schools were satisfied with the use of student achievement scores to assess performance (60 versus 69 percent), feedback received on performance (75 versus 80 percent), and recognition of accomplishments (60 versus 66 percent; Table V.1). Teachers in treatment and control schools responded similarly to the other satisfaction questions with one exception: treatment teachers were more satisfied with opportunities to earn extra pay (62 versus 54 percent).

Overall, no clear evidence exists that the impacts of pay-for-performance on teachers' satisfaction improved or worsened. Findings from Years 1 and 2 point to negative impacts on teachers' satisfaction with particular aspects of their jobs (although the satisfaction measures affected were not identical across years). Some initial negative impacts on teachers' satisfaction with their school environment and opportunities for advancement did not persist in the second year; however, by Year 2, there was new evidence of negative impacts on satisfaction with factors associated with how they were evaluated. Negative impacts on teachers' satisfaction with recognition of accomplishments persisted across both years.

The bonuses could affect some groups of teachers differently, so we examined impacts separately by subgroups. We separated teachers based on (1) grade-subject assignments (those in "tested" grades and subjects with annual accountability tests and those in "nontested" grades and subjects); and (2) experience levels (novice, mid-career, or late-career). These groupings stem from the hypothesis that teachers in tested grades and subjects could feel more pressure from the TIF program than teachers in nontested grades, because they could be evaluated on their own students' achievement growth or because the school's ability to receive a school-based award depended in part on their students' achievement. On the other hand, as shown in Chapter IV, teachers who were evaluated on their own students' achievement growth could earn higher bonuses than other teachers in the same districts. Separating teachers by their level of experience is of interest because teachers who had been teaching longer under a different evaluation and compensation system could have been less receptive to the new system.

The results of the subgroup analyses should be interpreted carefully. The impact estimate within each subgroup, which is based purely on the study's experimental design, captures the causal effect of pay-for-performance on outcomes within that subgroup.[52] However, a difference in impacts between two subgroups simply indicates whether impacts were larger or smaller in one subgroup than in another. It does not necessarily indicate whether the characteristic that distinguishes the two subgroups *caused* the difference in impacts, because characteristics other than the one being considered also might have differed between these subgroups. Nevertheless, because the subgroup analyses can identify the groups that respond most to pay-for-performance, they can inform best practices for designing or targeting future pay-for-performance programs.

**Pay-for-performance had the most negative effect on veteran teachers' satisfaction with factors associated with their evaluation and school environment.** For the three satisfaction measures on which pay-for-performance had an overall negative impact in Year 2, the negative impact

[52] By the second year of TIF implementation, it is possible that differences between treatment and control teachers within a subgroup are driven by differences in the composition of teachers in treatment and control schools. However, in Chapter VI, we found no evidence that pay-for-performance had an impact on the composition of the teaching workforce.

tended to be most pronounced among teachers with more than 15 years of experience (Table V.2).[53] Also, pay-for-performance increased less experienced teachers' satisfaction with school morale (Appendix E, Table E.5). In the other subgroups defined by experience levels or teaching assignments, impacts on satisfaction were not statistically significant (Appendix E, Table E.5).

**Table V.2. Impacts of Pay-for-Performance on Selected Teacher Satisfaction Measures for Teacher Subgroups, Year 2 (Percentage Points)**

| | Impacts on Whether Teachers Were "Somewhat" or "Very" Satisfied with… | | | | |
|---|---|---|---|---|---|
| Subgroup | Opportunities to Earn Extra Pay | Use of Student Achievement Scores to Measure Performance | Feedback on My Performance | Recognition of Accomplishments | **Number of Teachers[a]** |
| All Teachers (primary analysis) | 9* | -9* | -5* | -6* | **892-896** |
| Teaching Assignment | | | | | |
| (1) Tested grades and subjects | 8 | -10* | -6 | -7 | **484-487** |
| (2) Nontested grades and subjects | 10 | -7 | -3 | -6 | **407-410** |
| Difference between subgroups (1) - (2) | -2 | -4 | -3 | -2 | |
| Teacher Experience | | | | | |
| (1) Less than 5 years | 9 | 5 | 5 | 5 | **191-193** |
| (2) 5 to 15 years | 7 | -9* | -3 | -3 | **453-454** |
| (3) Greater than 15 years | 10 | -16* | -15* | -21* | **247-250** |
| Difference between subgroups (1) - (2) | 2 | 14 | 8 | 7 | |
| Difference between subgroups (3) - (2) | 3 | -7 | -12 | -18 | |

Source: Teacher survey, 2013.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

Among treatment teachers, we also looked at whether their attitudes vary by whether they received a bonus based on prior year's performance. This descriptive analysis can shed light on whether teachers may have had more favorable attitudes toward their job and TIF if they received a monetary reward for their performance. However, this analysis does not provide conclusive evidence about the effects of receiving bonuses on teachers' attitudes, because teachers who did and did not receive bonuses may have differed on many other characteristics that influenced their attitudes. Appendix E, Table E.6 presents the results on teachers' satisfaction by the second year of TIF implementation and shows no significant difference in any measure of teacher satisfaction between treatment teachers who had been awarded a bonus in Year 1 and those who had not.

**Principals in treatment schools were less satisfied than principals in control schools with the measures used to evaluate their performance.** In Year 2, the percentage of principals satisfied with aspects of their professional opportunities, evaluation system, and school environment ranged from 61 to 90 percent (Table V.3). As with teachers, however, principals in treatment schools were less satisfied than principals in control schools with some factors associated with how they were evaluated. Treatment principals were less satisfied than control principals with the use of observations

[53] Appendix E, Table E.5 presents findings for the other teacher satisfaction measures.

to assess skills (61 versus 85 percent) and the use of student achievement scores to assess performance (66 versus 82 percent). For other aspects of their professional opportunities and school environment, differences in satisfaction between treatment and control principals were generally negative but not statistically significant. In addition, treatment principals during the first year of TIF implementation reported being less satisfied than control principals with school morale (71 versus 87 percent); by Year 2, however, the impact was smaller and insignificant.

**Table V.3. Principals' Satisfaction with Professional Opportunities, Evaluation System, and School Environment (Percentages Who Are "Somewhat" or "Very" Satisfied)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Satisfaction Dimension | Treatment | Control | Impact | Treatment | Control | Impact |
| Opportunities for Pay and Development | | | | | | |
| Opportunities for professional advancement | — | — | — | 86 | 89 | -3 |
| Opportunities to enhance skills | 92 | 95 | -3 | 87 | 85 | 2 |
| Opportunities to earn extra pay | 72 | 66 | 6 | 63 | 64 | -1 |
| Factors Associated with Evaluation System | | | | | | |
| Use of observations to assess skills | — | — | — | 61 | 85 | -24* |
| Use of student achievement scores to assess performance | — | — | — | 66 | 82 | -16* |
| Feedback on my performance | 84 | 87 | -3 | 67 | 80 | -13 |
| School Environment | | | | | | |
| Recognition of accomplishments | 78 | 82 | -4 | 64 | 75 | -12 |
| Quality of interaction with colleagues | 90 | 97 | -7 | 86 | 90 | -4 |
| Colleagues' efforts | 93 | 98 | -5 | 90 | 85+ | 5 |
| School morale | 71 | 87 | -16* | 75 | 82 | -7 |
| **Number of Principals—Range[a]** | **63-64** | **59-61** | | **63-64** | **60-61** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

— is not available.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

## Educators' Attitudes Toward TIF

**Most teachers were glad to be participating in TIF, but teachers in treatment schools were less likely than teachers in control schools to be positive about TIF.** In both years of TIF implementation, approximately two-thirds of teachers were glad they were participating in TIF, and at least half felt TIF was fair (Table V.4). By Year 2, however, treatment teachers were more likely than control teachers to report that TIF reduced their freedom to teach the way they would like (40 versus 30 percent) and harmed the collaborative nature of teaching (29 versus 21 percent). In addition, treatment teachers were less likely than control teachers to believe student test scores measure what students learn (34 versus 41 percent). In Year 2, as in Year 1, pay-for-performance continued to cause a higher percentage of treatment teachers than control teachers to feel increased pressure to perform (65 versus 51 percent).

**Table V.4. Teachers' Attitudes Toward TIF Program (Percentages Who "Agree" or "Strongly Agree")**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Statement | Treatment | Control | Impact | Treatment | Control | Impact |
| Teachers who do the same job should receive the same pay | 57 | 58 | -1 | 61 | 66+ | -4 |
| Standardized student test scores in my district measure what students have learned | 35 | 33 | 2 | 34 | 41+ | -7* |
| My principal is a good judge of teacher talent | 67 | 73 | -6 | 74+ | 74 | 0 |
| I am glad that I am participating in the TIF program | 66 | 65 | 1 | 66 | 71+ | -5 |
| My job satisfaction has increased due to the TIF program | 28 | 33 | -5 | 38+ | 38 | 0 |
| I feel increased pressure to perform due to the TIF program | 65 | 53 | 11* | 65 | 51 | 14* |
| I have less freedom to teach the way I would like to teach due to the TIF program | 34 | 35 | 0 | 40 | 30 | 10* |
| The TIF program has harmed the collaborative nature of teaching | 23 | 24 | -1 | 29 | 21 | 8* |
| The TIF program has caused teachers to work more effectively | 49 | 46 | 3 | 50 | 56+ | -6 |
| The TIF program is fair | 53 | 58 | -5 | 54 | 59 | -5 |
| The process used to determine how bonuses are determined was adequately explained to me | 67 | 59 | 8* | 66 | 62 | 4 |
| **Number of Teachers—Range[a]** | **381-388** | **382-398** | | **397-440** | **383-442** | |

Source: Teacher survey, 2012 and 2013.

Notes: The statements in the table are identical to the language used in the survey. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

**Pay-for-performance had the most negative effect on veteran teachers' attitudes toward their TIF program.** Similar to the findings for satisfaction, we examined the impacts of pay-for-performance on teachers' attitudes toward TIF separately within subgroups defined by teaching assignment and level of experience. Pay-for-performance had a stronger, less favorable, impact on teachers with more than 15 years of experience (Appendix E, Table E.7). In most other subgroups and on most satisfaction measures, treatment teachers had less favorable attitudes than control teachers toward TIF; however, only in a few cases were those impacts statistically significant (Appendix E, Table E.7).

Among treatment teachers, we found little evidence in Year 2 that those who received a Year 1 performance bonus had more favorable attitudes toward TIF than those who did not. Attitudes toward most aspects of TIF did not differ between bonus recipients and nonrecipients (Table V.5). However, bonus recipients were *more* likely to think that TIF harmed the collaborative nature of teaching and less likely to believe that teachers who do the same job should receive the same pay.

**Table V.5. Attitudes of Teachers in Treatment Schools Toward TIF Program by Bonus Receipt, Year 2 (Percentages Who "Agree" or "Strongly Agree")**

| Statement | Received a Bonus After Year 1 | Did Not Receive a Bonus After Year 1 | Difference |
|---|---|---|---|
| Teachers who do the same job should receive the same pay | 38 | 60 | -22* |
| Standardized student test scores in my district measure what students have learned | 22 | 34 | -13 |
| My principal is a good judge of teacher talent | 76 | 72 | 4 |
| I am glad that I am participating in the TIF program | 56 | 60 | -3 |
| My job satisfaction has increased due to the TIF program | 42 | 29 | 13 |
| I feel increased pressure to perform due to the TIF program | 71 | 59 | 12 |
| I have less freedom to teach the way I would like to teach due to the TIF program | 48 | 43 | 5 |
| The TIF program has harmed the collaborative nature of teaching | 45 | 31 | 14* |
| The TIF program has caused teachers to work more effectively | 49 | 42 | 7 |
| The TIF program is fair | 48 | 45 | 3 |
| The process used to determine how bonuses are determined was adequately explained to me | 66 | 61 | 5 |
| **Number of Teachers—Range[a]** | **248-275** | **149-165** | |

Source: Teacher survey (2013) and educator administrative data.

Notes: Table is based on teachers in treatment schools. Pay-for-performance bonus receipt information comes from Year 1 educator administrative data. The statements in the table are identical to the language used in the survey. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**We found no clear evidence that principals' attitudes toward TIF differed between treatment and control schools.** We asked principals about their attitudes toward several aspects of TIF, such as the clarity with which the program had been communicated, the fairness of the evaluation system, and the program's effects on school staff. On most of these dimensions, differences in responses between treatment and control principals pointed toward less favorable attitudes among treatment principals. However, given the relatively small number of principals in the study, only large differences are likely to be statistically significant, and none of these differences met that threshold (Table V.6). Most treatment and control principals in Year 2 reported that the TIF program was clearly communicated to them (more than 90 percent), about half the principals reported that the evaluation system omitted important aspects of school administration that should be considered (54 percent of treatment principals and 48 percent of control principals), and over half (56 percent of treatment principals and 68 percent of control principals) reported that the TIF program contributed to greater collegiality and professionalism among the school staff.

**Table V.6. Principals' Attitudes Toward TIF Program (Percentage Who "Agree" or "Strongly Agree")**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Statement | Treatment | Control | Impact | Treatment | Control | Impact |
| The TIF program has been clearly communicated to me | 83 | 89 | -6 | 93 | 97 | -4 |
| This school has less chance of earning a bonus because of the characteristics of our student population | 22 | 20 | 3 | 38+ | 24 | 14 |
| The evaluation system omits important aspects of school administration that should be considered | 30 | 30 | 0 | 54+ | 48 | 6 |
| The TIF program contributes to greater collegiality and professionalism among the staff at this school | 49 | 55 | -6 | 56 | 68+ | -12 |
| Teachers at this school are more comfortable with frequent formal observations of their teaching because of the TIF program | 54 | 63 | -9 | 58 | 68 | -10 |
| Parents and the school community believe the TIF program is important | 39 | 48 | -8 | 50 | 43 | 7 |
| The TIF program is likely to continue for the foreseeable future | 85 | 87 | -2 | 71 | 73+ | -2 |
| I played an important role in implementing the TIF program at my school | 82 | 84 | -2 | 86 | 84 | 2 |
| **Number of Principals—Range[a]** | **62-65** | **60-64** | | **59-63** | **58-60** | |

Source: Principal survey, 2012 and 2013.

Note: The statements in the table are identical to the language used in the survey. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

## Impact of Pay-for-Performance on Educators' Behaviors

In this section, we present estimates of the impact of pay-for-performance on educators' behaviors. We discuss impacts on two types of behaviors, shown in the theory of change in Chapter I, which may shape the effectiveness of the teacher workforce: (1) how teachers use their time throughout the school day, and (2) principals' recruitment activities. The ways in which teachers allocate their time can influence how productive they are, and principals' recruitment of teachers can affect the composition and quality of their schools' staff.[54]

[54] In Appendix E, we report impacts on other principal behaviors that more indirectly affect teachers' productivity and retention, including principals' approaches to assigning teachers to grades and subjects and providing nonmonetary benefits to their teachers. We found no evidence that principals make decisions on teacher assignments or nonmonetary benefits differently in response to pay-for-performance.

## Teachers' Use of Time Throughout the School Day

We asked teachers to report how they spent their time in the most recent full week of teaching. In theory, pay-for-performance could motivate teachers to allocate more time to activities aimed at improving their performance ratings. For example, if efforts to improve performance ratings entail revamping lessons to be better aligned with state assessments, treatment teachers may decide to spend more time than control teachers on class preparation.

**By the second year of TIF implementation, pay-for-performance did not generally affect teachers' time on school-related activities.** On average, teachers in Year 2 reported working approximately 43 hours during school hours in the most recent full week of work—a significant increase from Year 1 of 5 hours of work for control teachers (Table V.7). Treatment and control teachers reported spending a similar amount of time on specific activities both during and outside school hours. The only exception is that treatment teachers spent about one hour less during the week on academic activities with students during nonschool hours than did control teachers.

**Table V.7. Teachers' Time Spent on School-Related Activities in the Most Recent Full Week (Average Hours)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Impact | Treatment | Control | Impact |
| Time Spent During School Hours on | | | | | | |
| Teaching students in the classroom, small groups, or individually | 27 | 26 | 1 | 28 | 28 | 0 |
| Supervising students in other activities | 4 | 4 | 0 | 4 | 4 | 0 |
| Preparation on your own (e.g., lessons, grading, assignment) | 6 | 7 | 0 | 8+ | 7 | 1 |
| Preparation and professional development with colleagues (e.g., common lesson planning, workshops, staff meetings, mentoring) | 3 | 3 | 0 | 4 | 4 | 0 |
| Other activities | 2 | 2 | 1 | 2 | 2 | 0 |
| Total hours during school hours (calculated) | 40 | 38 | 2 | 44 | 43+ | 0 |
| Time Spent During Nonschool Hours on | | | | | | |
| Academic-related activities with students | 2 | 2 | -1* | 3 | 4+ | -1* |
| Other activities with students | 1 | 1 | 0 | 1 | 1 | 0 |
| Preparation on your own | 9 | 8 | 0 | 10 | 8 | 2 |
| Preparation and professional development with colleagues | 2 | 3 | -1* | 3 | 3 | 0 |
| Other school-related activities | 1 | 1 | 0 | 1 | 1 | 0 |
| Total hours during nonschool hours (calculated) | 13 | 15 | -1 | 17+ | 17 | 1 |
| **Number of Teachers—Range[a]** | **321-393** | **315-402** | | **434-451** | **443-453** | |

Source: Teacher survey, 2012 and 2013.

Note: The categories in the table are identical to the language used in the survey. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

### Principals' Recruitment Efforts

To understand the possible impact of pay-for-performance on teacher recruitment, we asked principals whether and how they used TIF to recruit teachers to their school. Although all study principals might use opportunities offered through their TIF program to recruit teachers, we hypothesized that principals in schools that could offer pay-for-performance bonuses might recruit teachers differently because TIF offered teachers the possibility of earning higher bonuses in their schools than in control schools. In theory, being able to offer larger bonuses might help principals recruit more teachers and higher-performing teachers.

**Pay-for-performance had few effects on principals' approach to recruiting teachers.** When recruiting teachers, treatment and control principals generally reported emphasizing similar reasons for why teachers should work at their schools, with two exceptions (Table V.8). First, principals in treatment schools were more likely than principals in control schools to report using bonuses as a recruitment strategy in Year 1. Treatment principals also were more likely than control principals to report using bonuses to recruit teachers in Year 2, but the difference was not statistically significant. (Again, these findings may not be statistically significant because of the small number of principals in the study.) Second, treatment principals were more likely than control principals to emphasize the TIF program as a recruitment incentive during the first year of implementation (49 versus 29 percent), but they were not more likely to do so by the second year. The percentage of treatment principals emphasizing the TIF program was similar across years; by Year 2, however, 40 percent of control principals were also using this approach.

**Pay-for-performance had no impact on principals' success in hiring teachers.** Principals of treatment and control schools reported having similar recruitment experiences in terms of interviews per vacancy and acceptances per offer made. Based on the principals' reports, there were no statistically significant differences between treatment and control schools in the number of candidates interviewed per vacancy or the number of acceptances per job offer made in Years 1 and 2 (Table V.9).[55]

Although some differences between treatment and control principals' reports in Year 1 were consistent with higher teacher turnover in control schools—more vacancies, interviews conducted, and offers made—evidence from administrative data, presented later in this report, does not point to either a positive or negative impact of pay-for-performance on teacher retention (Appendix F, Table F.5).[56]

[55] Most principals in both treatment and control schools reported having input in hiring decisions. Fewer than 3 percent of principals reported having little or no input in hiring teachers at their school (Appendix E, Table E.8).

[56] Appendix E, Tables E.9 and E.10 show that principals in treatment and control schools also reported using similar criteria for assigning teachers to grades or subject areas, and were similar in their use of nonmonetary benefits to reward teachers.

**Table V.8. Incentives Used to Recruit Teachers (Percentages Who Reported They Were "Always" or "Often" Used)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Incentives | Treatment | Control | Impact | Treatment | Control | Impact |
| Salary | 23 | 22 | 0 | 21 | 22 | 0 |
| Opportunities to earn performance-based pay | 27 | 14 | 12* | 33 | 17 | 16 |
| Opportunities for career advancement | 25 | 21 | 4 | 27 | 28 | -1 |
| Opportunities for professional development | 62 | 62 | 0 | 66 | 57 | 9 |
| The level of teacher involvement in school decision making | 49 | 57 | -8 | 53 | 52 | 0 |
| Collegiality of teaching staff | 78 | 86 | -8 | 79 | 88 | -9 |
| The school culture and/or educational philosophy | 86 | 86 | 0 | 81 | 92 | -11 |
| The school's reputation | 74 | 72 | 2 | 64 | 77 | -12 |
| The school's location or neighborhood | 38 | 41 | -2 | 29 | 28 | 1 |
| The level of student achievement at the school | 52 | 51 | 2 | 45 | 44 | 1 |
| The TIF program | 49 | 29 | 20* | 45 | 40 | 5 |
| **Number of Principals—Range[a]** | **61-64** | **62-64** | | **61-64** | **60-61** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table V.9. Teaching Vacancies and Hiring Experiences (Averages Unless Otherwise Noted)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Impact | Treatment | Control | Impact |
| Classroom with teacher vacancies | 3 | 4 | -2* | 4 | 5 | -1 |
| Applications school reviewed for positions | 28 | 28 | -1 | 33 | 33 | -1 |
| Applicants school interviewed | 10 | 14 | -4* | 11 | 17+ | -6* |
| Offers school made | 3 | 5 | -2* | 4 | 5 | -1 |
| Offers that were accepted | 3 | 5 | -2* | 4 | 4 | -1 |
| Interview ratio (interviewed applicants divided by classroom vacancies) (percentage) | 26 | 32 | -6 | 36 | 35 | 1 |
| Acceptance ratio (offers accepted divided by offers made) (percentage) | 75 | 81 | -5 | 81 | 84 | -3 |
| **Number of Principals—Range[a]** | **58-63** | **60-63** | | **61-64** | **58-61** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

## Summary

The ways in which pay-for-performance programs affect educators' attitudes and behaviors can shape how pay-for-performance affects student outcomes. The goal of pay-for-performance is to improve student achievement by motivating educators to improve their performance and by attracting and retaining more effective teachers. However, if the presence of pay-for-performance discourages useful collaboration, lowers morale, or makes a school less appealing to effective educators, it may not accomplish this goal.

The findings from this chapter are mixed about whether the changes in attitudes resulting from pay-for-performance would enhance or hinder educators' effectiveness. Most teachers and principals reported being satisfied with key aspects of their job and TIF program. However, pay-for-performance made teachers less satisfied with factors associated with how they were evaluated, their school environment, and their TIF program, but increased teachers' satisfaction with their opportunity to earn extra pay. These effects could have offset each other in shaping educators' motivation to work more effectively or to work in schools that offer performance bonuses. In addition, given our findings that veteran teachers responded least favorably to pay-for-performance, districts may find that acceptance of pay-for-performance programs will improve over time as veteran teachers retire. Until then, districts that plan to adopt this policy may benefit from tailoring their communication to teachers at different experience levels.

# VI. IMPACTS OF PAY-FOR-PERFORMANCE ON EDUCATOR EFFECTIVENESS AND STUDENT ACHIEVEMENT

A central objective of the TIF grants is to improve student achievement in high-need schools by increasing the effectiveness of the educators working in those schools. Our evaluation was designed to rigorously assess whether the pay-for-performance component of grantees' TIF programs accomplished this goal. In this chapter, we present findings on whether pay-for-performance led to changes in educator effectiveness and student achievement after one and two years of TIF implementation.

**Key Findings After Two Years**

- **Pay-for-performance led to teachers and principals earning higher effectiveness ratings based on student achievement growth in their schools, but did not affect ratings based on observations of their classroom or school practices.**
- **Pay-for-performance did not enable schools to retain or attract more higher-performing teachers.**
- **Pay-for-performance led to more higher-performing principals staying in their schools and more lower-performing principals leaving their schools.**
- **Pay-for-performance had small, positive impacts on students' reading achievement that were equivalent to about three weeks of additional learning. Pay-for-performance had similar, but insignificant, impacts on students' math achievement.**
- **The impacts of pay-for-performance on student achievement differed among districts, but differences in impacts were not related to differences in key program characteristics measured by this study.**

As shown in the theory of change from Chapter I, a main principle of TIF is that increasing educator effectiveness is the key to improving student achievement. Therefore, this chapter begins by reporting the impacts of pay-for-performance on educator effectiveness. Pay-for-performance could lead to greater educator effectiveness by either enabling schools to attract and retain more effective educators or motivating educators to improve their effectiveness. Therefore, in the second section of this chapter, we examine specifically whether pay-for-performance led to changes in the retention and recruitment of effective educators. Throughout this chapter, we measure educator effectiveness with the performance ratings that educators received from their districts. Those ratings were largely based on measures of student achievement growth in classrooms and schools, and observations of classroom or school practices. Because those ratings determined performance bonus amounts, pay-for-performance was designed to motivate educators to improve their performance on those measures. However, those measures might not capture all aspects of educator performance that matter for student achievement. Therefore, in the last section of this chapter, we directly examine whether pay-for-performance bonuses led to improved student achievement on reading and math assessments.

Our analyses in this chapter compare the outcomes of educators and students in treatment schools with those of educators and students in control schools. Educators in treatment schools were eligible for pay-for-performance bonuses and educators in control schools were not. Because both treatment and control schools offered all the other required components of the TIF program, any differences in outcomes between treatment and control schools can be attributed to the impact of

pay-for-performance.[57] Data for this chapter come from districts' administrative records on educators and students.

The chapter is based on 10 evaluation districts that completed two years of TIF implementation during the period covered by this report. We refer to the first and second years of implementation, 2011–2012 and 2012–2013, as Years 1 and 2.[58] Examining impacts in both years provided an opportunity to see whether impacts evolved over time. For example, impacts could have been larger in Year 2 than Year 1 for several reasons. Educators' understanding of their evaluation measures and bonus eligibility increased over time (see Chapter IV), and educators could have also been more motivated to improve in Year 2 after seeing the first round of performance bonuses, which were awarded after Year 1 was completed. Moreover, it could have taken time for educators to change their practices or decisions on where to work in response to the opportunity to earn performance bonuses.

## Impact of Pay-For-Performance on Educator Performance Ratings

Pay-for-performance can increase or decrease educator effectiveness—or simply not matter. Rewarding high-performing educators with performance bonuses could increase educator effectiveness if it motivates educators to improve or encourages high-performing educators to work in schools that offer those bonuses. On the other hand, pay-for-performance could have no effect on educator effectiveness if educators are not motivated by monetary incentives, do not find the incentives large enough to change their practices, or cannot identify ways to improve their performance. It might even lead to a less effective educator workforce if it discourages cooperation, lowers morale, or causes more effective educators to leave schools. In fact, evidence from Chapter V indicated that pay-for-performance lowered teachers' satisfaction with factors associated with how they were evaluated and school environment, but raised their satisfaction with opportunities to earn extra pay, raising the question of whether any of these changes in satisfaction influenced their effectiveness. In this section, we report the impacts of pay-for-performance on educator effectiveness, capturing the combined influence of all of these possible effects.

The measures of educator effectiveness for this analysis were those that districts used to evaluate educators' performance as part of their TIF programs. Districts had to evaluate teachers and principals based on student achievement growth and at least two observations of classroom or school practices. However, districts had flexibility in how they implemented this requirement. For example, they could choose to evaluate teachers based on the achievement growth of the teachers' own students (classroom achievement growth), all students in the same grade, all students in the school (school achievement growth), or some combination of these measures. The advantage of examining these measures is that TIF programs provided an incentive for educators to improve their performance on

[57] The final section of Appendix F provides supplemental information on the quality of the analyses conducted for this chapter (see Tables F.18 and F.19). In addition, as discussed in Chapter IV, some educators in the study schools misunderstood their eligibility for pay-for-performance or the potential amounts they could earn. The impacts reported in this chapter reflect the impact of pay-for-performance given educators' actual beliefs. This study was not designed to assess the impacts of pay-for-performance bonuses if all educators correctly understood their eligibility or the amount they could earn in a bonus.

[58] As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts examined in this chapter, whose schools were randomly assigned in spring and summer 2011, were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort. In Appendix F, we present Year 1 impacts on educator effectiveness and student achievement for Cohorts 1 and 2 together—that is, findings from 2011–2012 for Cohort 1 and from 2012–2013 for Cohort 2.

those measures. The disadvantage is that educator effectiveness was not measured in a consistent way across districts (see Chapter IV). For example, six evaluation districts used growth measures provided by the state and four districts used models developed by private vendors. Districts also used a variety of observation rubrics.

We examined the impact of pay-for-performance on four measures of educator effectiveness obtained from district administrative records: (1) school achievement growth ratings, which were used to evaluate teachers and principals; (2) classroom achievement growth ratings for teachers; (3) classroom observation ratings for teachers; and (4) observation ratings for principals. Each of these performance measures placed educators into three to five performance categories—such as effective or highly effective—or on a numeric scale in which an increase of one point was similar to advancing one performance level. To express ratings from different districts on a common scale, we expressed each rating as a score on a 1-to-4 rating scale, with 1 being the lowest and 4 being the highest possible rating an educator could receive on the district's measure of performance (see Appendix B for details). Thus, an increase from 3 to 4 on the rating scale can roughly be interpreted as a change from being classified as effective to being classified as highly effective.

We examined each performance measure separately for two reasons. First, the different measures may capture different aspects of effectiveness. For example, classroom observations could have identified aspects of teachers' instruction that mattered for classroom climate but not for students' math or reading achievement. Second, as discussed in Chapter IV, districts awarded separate bonuses for each performance measure, so educators could have focused on improving their performance on the measures that they could influence most easily or that were tied to the largest bonuses.

The findings below capture the impacts of pay-for-performance bonuses on average educator performance ratings in schools that offered those bonuses. As we discuss later in this chapter, average ratings in schools could change for a variety of reasons, including improvements in educators' practices and the hiring or departure of higher- or lower-performing educators. For simplicity, we describe the findings as impacts on teachers' or principals' ratings, but these statements are shorthand for impacts on the average educator performance ratings of schools.

**Districts' Measures of Student Achievement Growth in Classrooms and Schools**

The two most common student achievement growth measures that districts used to evaluate educators were those that measured achievement growth of all students in a school and in teachers' specific classrooms (see Chapter IV). In theory, school achievement growth combines the contributions of all staff at a school, so impacts on school achievement growth might reflect how teachers, principals, or other school staff responded to pay-for-performance. In 6 of the 10 districts, some teachers were also evaluated on student achievement growth in their own classrooms. In those districts, teachers who received classroom achievement growth ratings were typically those who taught grades and subjects that were tested using annual state assessments.

**Pay-for-performance led to teachers and principals earning higher effectiveness ratings based on the achievement growth of all students in their schools.** On a 1-to-4 rating scale, educators in treatment schools had school achievement growth ratings that were 0.34 points higher than those of educators in control schools in Year 1, and 0.25 points higher in Year 2 (Table VI.1).[59,60]

**Among teachers who were evaluated on student achievement growth in their own classrooms, pay-for-performance led to teachers earning higher classroom achievement growth ratings in Year 1, but not in Year 2.** In Year 1, teachers in treatment schools had classroom achievement growth ratings that were 0.18 points higher than those of teachers in control schools (Table VI.1). In Year 2, however, the 0.04 point difference between these groups of teachers was not statistically significant.

**Table VI.1. Student Achievement Growth Ratings (Points on 1-to-4 Scale)**

| Performance Measure and Year | Treatment | Control | Impact | *p*-value | **Number of Teachers** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **School Achievement Growth** | | | | | | |
| Ratings in Year 1 | 2.59 | 2.25 | 0.34* | 0.046 | **NA** | **124[a]** |
| Ratings in Year 2 | 2.46 | 2.21 | 0.25* | 0.047 | **NA** | **131** |
| **Classroom Achievement Growth[b]** | | | | | | |
| Ratings in Year 1 | 2.26 | 2.08 | 0.18* | 0.033 | **1,093** | **73** |
| Ratings in Year 2 | 2.20 | 2.16 | 0.04 | 0.459 | **1,342** | **73** |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]School achievement growth ratings for one district in Year 1 were not included because they could not be converted to a 1-to-4 rating scale.

[b]Classroom achievement growth ratings are available only for the six districts that evaluated teachers based on classroom achievement growth.

*Impact is statistically significant at the .05 level, two-tailed test.

NA is not applicable.

The finding that pay-for-performance did not raise classroom achievement growth ratings in Year 2 differs from our earlier finding that pay-for-performance raised school achievement growth ratings. Both types of ratings ought to have reflected the achievement growth of students tested by state assessments. Nevertheless, there are some reasons why the two types of ratings need not have come to the same conclusions. First, whereas findings for school achievement growth were based on all districts, findings for classroom achievement growth were based only on six districts that used this measure.[61] Second, achievement growth by individual students need not have factored into school and classroom achievement growth ratings in the same way. For example, if many students at a school experienced greater achievement growth but each teacher taught only a few of those students, a district

[59] Appendix F, Tables F.1 and F.2 show findings from alternative ways of estimating impacts on school achievement growth ratings and classroom observation ratings.

[60] The estimated impact on school achievement growth ratings in Year 1 was not statistically significant when Cohort 2 schools were included in the analysis (Appendix F, Table F.3).

[61] In fact, within those six districts, impacts on school achievement growth ratings were smaller than the overall impacts and were not significant in either year (results not shown).

could have found sufficient evidence to conclude that the entire school was more effective but insufficient evidence to conclude that any one teacher was more effective.

### Observation Ratings for Teachers and Principals

In all districts, both teachers and principals received ratings based on formal observations of their practices. Trained observers rated teachers on their classroom practices and rated principals on the practices they implemented in their schools.

**Pay-for-performance had no impact on the observation ratings that either teachers or principals earned.** In Years 1 and 2, treatment and control teachers earned similar classroom observation ratings (Table VI.2).[62] Likewise, in both years, there were no statistically significant differences between observation ratings for principals in treatment and control schools.

**Table VI.2. Observation Ratings for Teachers and Principals (Points on 1-to-4 Scale)**

| Performance Measure and Year | Treatment | Control | Impact | *p*-value | Number of Educators | Number of Schools |
|---|---|---|---|---|---|---|
| **Teachers' Classroom Observation Ratings** | | | | | | |
| Ratings in Year 1 | 2.94 | 2.91 | 0.03 | 0.243 | **3,625** | **132** |
| Ratings in Year 2 | 3.02 | 2.97 | 0.05 | 0.070 | **3,628** | **132** |
| **Observation Ratings for Principals** | | | | | | |
| Ratings in Year 1 | 3.08 | 3.18 | -0.10 | 0.197 | **105** | **105[a]** |
| Ratings in Year 2 | 3.16 | 3.03 | 0.13 | 0.184 | **118** | **117** |

Source: Educator administrative data.

Notes: None of the impacts was statistically significant at the .05 level. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]One district did not provide observation ratings for principals in Year 1.

## Impact of Pay-for-Performance on the Retention and Recruitment of Effective Educators

The findings in the previous section indicate that pay-for-performance increased schools' effectiveness ratings as measured by student achievement growth. The theory of change, presented in Chapter I, suggests two possible explanations for this positive impact. First, pay-for-performance could have caused educators to improve. For example, the prospect of earning performance bonuses could have motivated educators to enroll in targeted professional development to learn about effective practices or collaborate more effectively with their colleagues. Second, pay-for-performance could have led more higher-performing educators to choose to work at these schools. Data on educators' school assignments combined with information on their effectiveness enabled us to examine the second explanation directly. This section reports findings on whether pay-for-performance caused

[62] The finding that pay-for-performance had no impact on classroom observation ratings is consistent when Cohort 2 schools were included in the analysis for Year 1 (Appendix F, Table F.4).

changes in staffing that resulted in more effective educators working at schools with pay-for-performance.[63]

There are two ways in which pay-for-performance could reshape schools' staff to include more higher-performing educators. First, pay-for-performance could enable schools to retain more higher-performing educators or encourage more lower-performing educators to leave. For example, higher-performing educators might be more likely to feel that their contributions are recognized as a result of receiving performance bonuses. Lower-performing educators might become discouraged if their colleagues receive bonuses and they do not, and therefore choose to leave. Second, pay-for-performance could enable schools to recruit more higher-performing educators to fill vacancies. As discussed in Chapter V, one-third of principals at treatment schools reported using pay-for-performance as a recruitment tool for hiring teachers, which suggests that these principals could have believed the offer of performance bonuses would attract better teachers.

The extent of educator turnover at a school determines how much staffing changes can affect the overall effectiveness of the school's educators. For example, if a large school had only one teacher depart each year, then overall effectiveness would change little if the departing teacher was the worst teacher rather than the best. Likewise, the effectiveness of the departing teacher's replacement would have little influence on overall effectiveness. In the study schools, about one-fifth of teachers departed from one year to the next, and one-third of teachers departed over a two-year period (Appendix F, Table F.5). Likewise, about one-fifth to one-fourth of principals departed from one year to the next, and two-fifths of principals departed over a two-year period (Appendix F, Table F.6). Therefore, although many educators were retained, there was also plenty of turnover, leaving the potential for staffing changes to be an important way of shaping educator effectiveness.[64]

**Retention of More Effective Teachers**

To assess whether pay-for-performance enabled schools to retain more effective teachers, we examined differences between treatment and control schools in the effectiveness of *both* teachers who stayed at and those who left their schools. If pay-for-performance led to the retention of more effective teachers, the performance ratings of retained teachers ought to have been higher in treatment schools than control schools. However, because pay-for-performance did not affect the overall percentage of teachers who stayed in their schools (Appendix F, Table F.5), any staffing changes that caused more higher-performing teachers to stay would have also caused more lower-performing teachers to leave. In this case, the performance ratings of departing teachers should have been *lower* in treatment schools than control schools.

Among teachers working in study schools in Year 1, we used performance ratings in Year 1 to measure the effectiveness of teachers who subsequently chose to stay at or leave their schools. We

[63] As explained in Chapter II, the study design required that half of the participating schools within a district would implement pay-for-performance bonuses and the other half would not. This design is likely to have led to larger mobility impacts than if pay-for-performance had been implemented district-wide. Pay-for-performance could have also altered other characteristics of the schools' staff, such as their demographic and professional characteristics. However, we found no evidence that pay-for-performance led to changes in those characteristics (Appendix F, Table F.7).

[64] The analyses in this section examine whether educators stayed at or left their original treatment or control school. Although it is possible that educators may have left their original school to go to another study school, in practice this was very rare. For example, between Years 1 and 3, only one percent of treatment teachers moved to another treatment school and two percent moved to a control school. In contrast, 31 percent of teachers left their original school for a school or position outside of the study. Since moving from one study school to another was very rare, we did not explicitly take into account this type of move.

focused on the two measures of individual teachers' performance—classroom observations and classroom achievement growth—because measures of school performance could not distinguish more- and less-effective teachers in the same school. We classified teachers in Year 1 as having subsequently stayed or left based on whether they taught in the same school from Year 1 to the fall of Year 3. We did not classify teachers based on where they worked in Year 2, because teachers might not have had time to fully take into account the ratings they earned in Year 1 when deciding where to work in Year 2. Few districts (only 3 of 10) awarded any performance bonuses based on Year 1 ratings before the start of the subsequent school year (see Chapter IV), so in most districts teachers were finalizing job decisions for Year 2 before seeing the bonuses they did or did not earn. For example, a treatment teacher who received a low rating in Year 1 might not have felt discouraged from staying until she realized that her colleagues received bonuses in the fall of Year 2 whereas she did not. Teachers' decisions on where to work in Year 3 might have more fully reflected the rewards or pressures resulting from their Year 1 performance ratings.

**Pay-for-performance did not enable schools to retain more higher-performing teachers.** In treatment and control schools, the teachers who remained from Years 1 to 3 did not significantly differ in effectiveness, as measured by either their classroom observation ratings or classroom achievement growth ratings from Year 1 (Figure VI.1). Likewise, there were no significant differences in performance ratings between teachers who left treatment schools and those who left control schools.[65] Therefore, the positive impacts of pay-for-performance on school achievement growth were not due to the retention of teachers regarded as more effective by their districts.

### Recruitment of More Effective Teachers

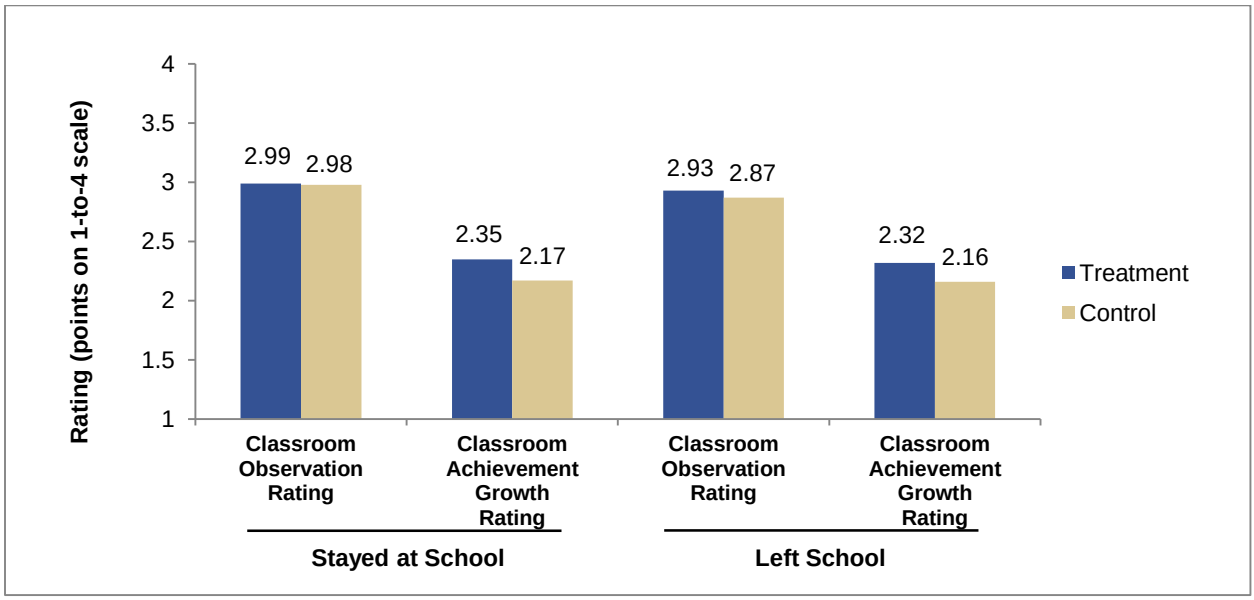
To examine whether pay-for-performance enabled schools to recruit more effective teachers to fill vacancies, we compared the Year 2 performance ratings of teachers in treatment and control schools who were new to their schools in that year. We focused on new recruits in Year 2 because teachers' decisions on where to work in Year 2 could have been shaped by districts' and schools' efforts in Year 1 to make teachers aware of the TIF program.

**Pay-for-performance did not enable schools to fill vacancies with more higher-performing teachers.** On both classroom observations and classroom achievement growth, newly hired teachers at treatment and control schools earned similar ratings in Year 2 (Figure VI.2).[66] Therefore, the effectiveness of newly hired teachers does not explain why treatment schools had higher achievement growth ratings than control schools in Year 2.

[65] Findings were similar when we examined the Year 1 performance ratings of teachers who stayed at and left their schools between Years 1 and 2, and the Year 2 performance ratings of teachers who stayed at and left their schools between Years 2 and 3 (Appendix F, Table F.8).

[66] Given that schools were randomly assigned to the treatment and control groups in the spring and summer before Year 1, it is possible that pay-for-performance could have enabled schools to recruit better teachers for Year 1. However, we found no evidence to support this possibility. Among teachers who were new to their schools in Year 1, performance ratings in Year 1 were similar for treatment and control teachers (Appendix F, Table F.10).

**Figure VI.1. Year 1 Performance Ratings of Teachers Who Stayed at and Left Their Schools Between Years 1 and 3 (Points on 1-to-4 Scale)**

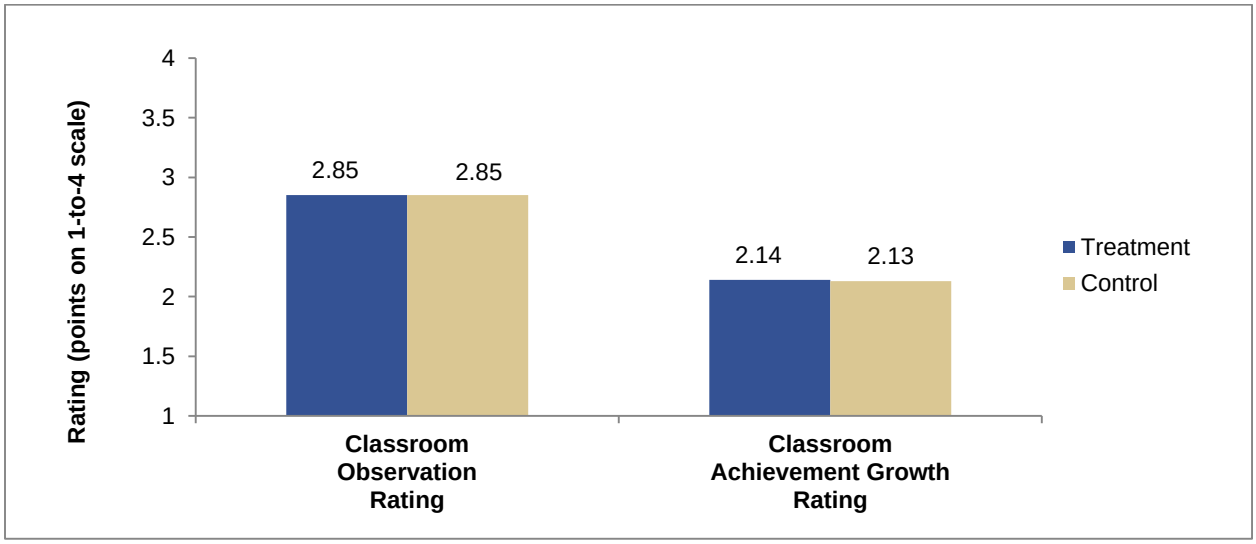



Source: Educator administrative data (N = 2,460 teachers for classroom observation ratings of teachers who stayed; N = 702 teachers for classroom achievement growth ratings of teachers who stayed; N = 1,165 teachers for classroom observation ratings of teachers who left; and N = 391 teachers for classroom achievement growth ratings of teachers who left).

Note: None of the differences between teachers in treatment and control schools was statistically significant at the .05 level.

Figure reads: Among teachers who stayed at their schools between Years 1 and 3, those in treatment schools earned an average classroom observation rating of 2.99 points in Year 1, and those in control schools earned an average classroom observation rating of 2.98 points in Year 1.

**Figure VI.2. Year 2 Performance Ratings of Teachers Who Were New to Their Schools in Year 2 (Points on 1-to-4 Scale)**




Source: Educator administrative data (N = 781 teachers for classroom observation rating and N = 351 teachers for classroom achievement growth rating).

Note: None of the differences between teachers in treatment and control schools was statistically significant at the .05 level.

Figure reads: Among teachers who were new to their schools in Year 2, the average classroom observation rating in Year 2 was 2.85 points in both treatment and control schools.

### Retention of More Effective Principals

Because pay-for-performance bonuses were also awarded to principals in treatment schools with high performance ratings, these principals could have been more motivated to stay in their schools than their counterparts in control schools. We compared the performance ratings of principals who stayed in treatment schools and those who stayed in control schools to assess whether pay-for-performance caused more higher-performing principals to stay. We also compared the performance ratings of principals who left treatment schools and those who left control schools to assess whether pay-for-performance caused more lower-performing principals to leave.

**Pay-for-performance led to more higher-performing principals staying in their schools and more lower-performing principals leaving their schools.** Evidence from school achievement growth ratings, but not observation ratings, indicates that more higher-performing principals stayed in treatment schools than control schools. Among principals who stayed in their schools from Years 1 to 3, school achievement growth ratings from Year 1 were higher among treatment principals than control principals by 0.6 points—a difference that was more than halfway between two performance levels on the four-level rating scale (Figure VI.3). However, the two groups earned similar observation ratings in Year 1.

On the other hand, evidence from observation ratings, but not school achievement growth ratings, indicates that more lower-performing principals left treatment schools than left control schools. Observation ratings in Year 1 were lower, by 0.3 points, among principals who left treatment schools than those who left control schools, but school achievement growth ratings were not significantly different (Figure VI.3).[67]

### Recruitment of More Effective Principals

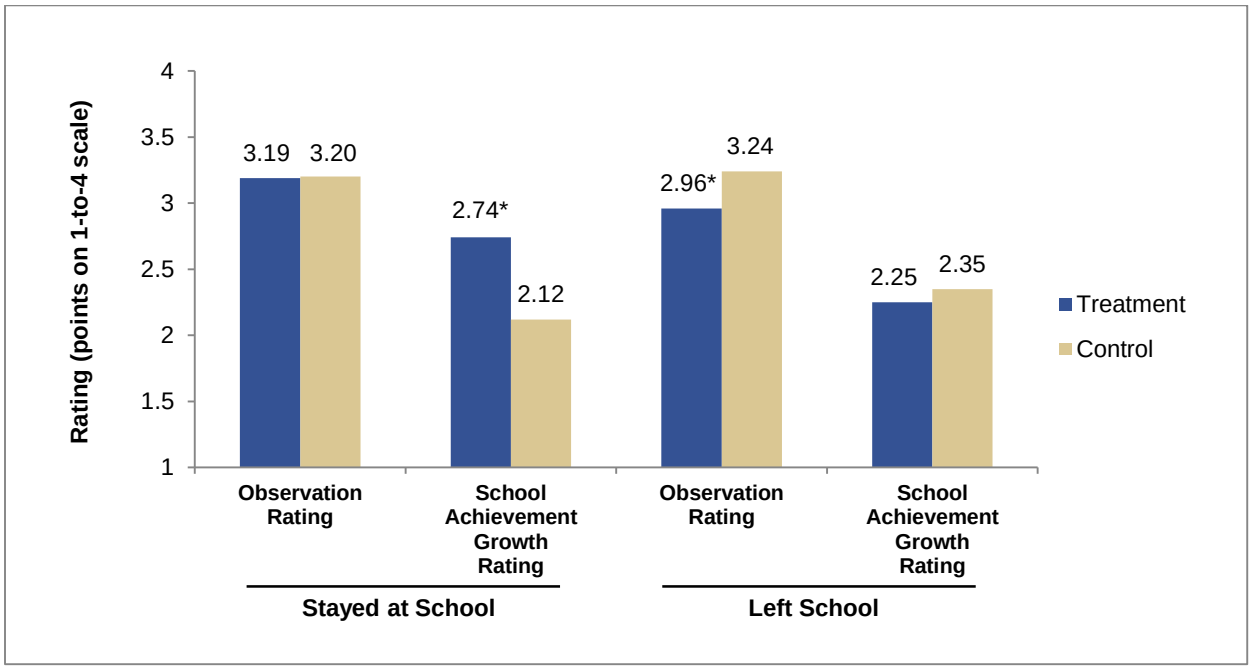
Beside encouraging higher-performing principals to stay in their schools, pay-for-performance could have also motivated more higher-performing principals to enter schools where they would be eligible for performance bonuses. To assess this possibility, we compared the performance ratings of principals who were newly hired to lead treatment and control schools in Year 2. Given the small number of principals, only relatively large differences would result in a statistically significant difference.

**Pay-for-performance did not lead to more higher-performing principals being hired at schools that offered performance bonuses.** On both observations and school achievement growth in Year 2, principals who were newly hired to lead treatment schools earned higher ratings than those who were newly hired to lead control schools, but these differences were not statistically significant (Figure VI.4).[68]

[67] We also examined the Year 1 performance ratings of principals who stayed at and left their schools between Years 1 and 2, and the Year 2 performance ratings of principals who stayed at and left their schools between Years 2 and 3 (Appendix F, Table F.9). As in the main findings, school achievement growth ratings were consistently higher among principals who stayed at treatment schools than those who stayed at control schools. However, we did not find statistically significant differences in performance ratings between principals who left treatment schools and those who left control schools.

[68] When examining newly hired principals in Year 1, we found that treatment principals earned higher school achievement growth ratings in that year than control principals did (Appendix F, Table F.11). Although this finding could suggest that pay-for-performance led to the recruitment of more effective principals in Year 1, it is unclear whether schools'

**Figure VI.3. Year 1 Performance Ratings of Principals Who Stayed at and Left Their Schools Between Years 1 and 3 (Points on 1-to-4 Scale)**



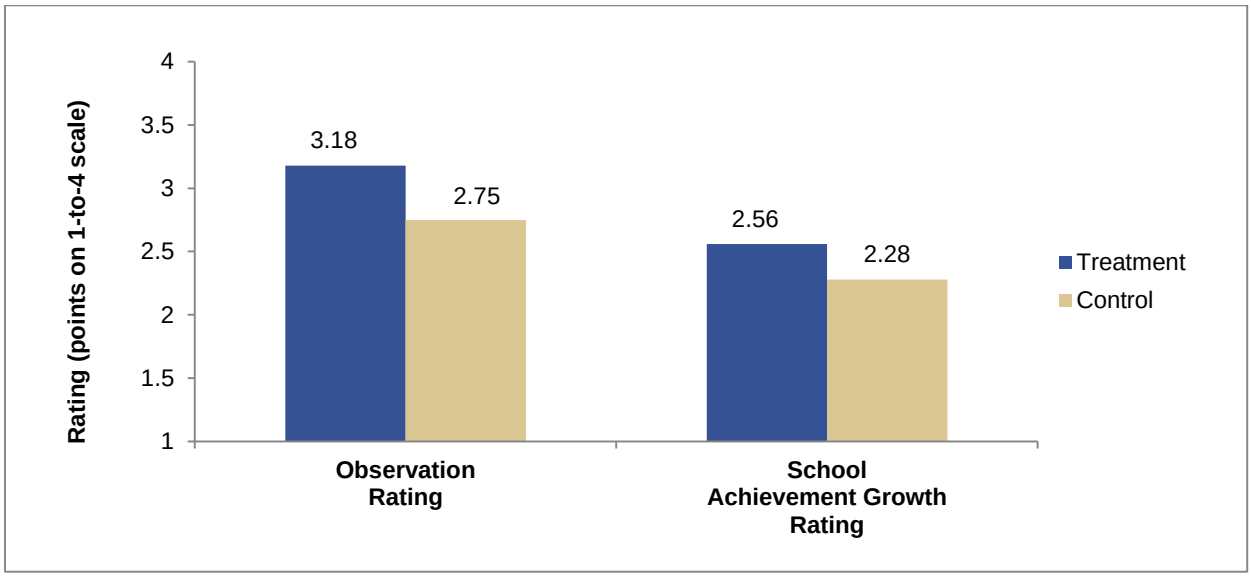

Source: Educator administrative data (N = 67 principals for observation ratings of principals who stayed; N = 73 principals for school achievement growth ratings of principals who stayed; N = 38 principals for observation ratings of principals who left; and N = 54 principals for school achievement growth ratings of principals who left).

Figure reads: Among principals who stayed at their schools between Years 1 and 3, those in treatment schools earned an average observation rating of 3.19 points in Year 1, and those in control schools earned an average observation rating of 3.20 points in Year 1.

*Difference between principals of treatment and control schools is statistically significant at the .05 level, two-tailed test.

eligibility for pay-for-performance—determined in the spring and summer before Year 1—would have been known to prospective principals at the time of hire. An alternative explanation for this finding is that pay-for-performance could have motivated newly hired principals in treatment schools to work more effectively than their counterparts in control schools in Year 1. A third explanation is that positive impacts on school achievement growth ratings could have also reflected improvements by teachers and other school staff—not just principals.

**Figure VI.4. Year 2 Performance Ratings of Principals Who Were New to Their Schools in Year 2 (Points on 1-to-4 Scale)**




Source: Educator administrative data (N = 19 principals for observation rating and N = 30 principals for school achievement growth rating).

Note: None of the differences between principals of treatment and control schools was statistically significant at the .05 level.

Figure reads: Among principals who were new to their schools in Year 2, those in treatment schools earned an average observation rating of 3.18 points in Year 2, and those in control schools earned an average observation rating of 2.75 points in Year 2.

## Impact of Pay-For-Performance on Student Achievement

TIF grants were designed to improve student achievement by increasing the effectiveness of teachers and principals. Although this chapter has shown that the pay-for-performance component of TIF increased educators' ratings on some performance measures, this does not necessarily translate into higher achievement for students. There is no guarantee that the performance measures used by TIF districts accurately captured aspects of teaching or leadership quality that might be important for student achievement. Therefore, in this section, we directly examine the impact of pay-for-performance on student achievement in the study schools, using administrative data on students' reading and math scores from state assessments.

Although higher performance ratings might not necessarily lead to improved student achievement, the findings presented earlier strongly suggest that pay-for-performance ought to have raised student achievement. The ratings that increased as a result of pay-for-performance were those that measured student achievement growth in schools. Districts calculated those ratings based, at least in part, on the same test scores that we collected for the analysis in this section.

Nevertheless, there are two key differences between the analysis in this section and the earlier analysis of ratings based on student achievement growth in schools. First, this analysis used the same method for all districts to analyze and compare student achievement in treatment and control schools. This is important because, as discussed in Chapter IV, districts used different methods both to construct their own measures of school achievement growth from test score data and then to convert those measures into ratings. Second, this analysis enabled us to examine the impacts of pay-for-performance separately on math and reading achievement. School achievement growth ratings generally combined achievement data from math and reading—and, in some cases, data from other subjects.

As discussed in Chapter II, we standardized test scores from different states and grades into $z$-scores, which reflected how well each student scored when compared with the average student in his or her state and grade. The findings after Year 1 show the impact of pay-for-performance on schools' average student achievement after the first year of implementation, when the program was new and educators had not yet received bonuses they earned based on their performance. The findings after Year 2 show the cumulative impact on schools' average student achievement after two years of implementation. For simplicity, we describe the findings as impacts on students' achievement, but these statements are shorthand for impacts on the average student achievement of schools.

**Pay-for-performance had small, positive impacts on students' reading achievement. It had similar, but insignificant, impacts on students' math achievement.** Students in treatment schools scored 0.03 standard deviations higher on reading assessments in Years 1 and 2 than students in control schools (Table VI.3). In math, differences in student achievement between treatment and control schools were also positive and similar in magnitude as those in reading, but not statistically significant ($p$-value = 0.34 in Year 1 and $p$-value = 0.07 in Year 2). As the negative $z$-scores indicate, the average achievement of students in both treatment and control schools was below the statewide mean, reflecting the fact that study schools were low-performing schools.

**Table VI.3. Student Achievement in Math and Reading (Student *z*-score units)**

| Year and Subject | Treatment | Control | Impact | *p*-value | **Number of Students** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Year 1** | | | | | | |
| Math | -0.43 | -0.45 | 0.02 | 0.335 | **40,852** | **132** |
| Reading | -0.37 | -0.40 | 0.03* | 0.040 | **40,576** | **132** |
| **Year 2** | | | | | | |
| Math | -0.39 | -0.43 | 0.04 | 0.068 | **40,709** | **132** |
| Reading | -0.36 | -0.39 | 0.03* | 0.026 | **40,391** | **132** |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

*Impact is statistically significant at the .05 level, two-tailed test.

There are several ways to interpret the magnitudes of the impacts on student achievement. First, the impacts can be expressed as a difference in percentiles of achievement. In Year 2, the average student in a control school earned a reading $z$-score of -0.39, placing that student at approximately the 35th percentile of student achievement statewide.[69] The average student in a treatment school earned a $z$-score of -0.36, representing approximately the 36th percentile—a gain of 1 percentile point. Similarly, impacts on reading achievement after Year 1 lifted the average student in these schools from the 34th to the 36th percentile. Impacts on math achievement, although not statistically significant, kept the average student at about the 33rd percentile after Year 1 and moved the average student from 33rd to the 35th percentile after Year 2.

The impacts can also be compared with the average one-year gain in achievement for students in grades 3 through 8 on nationally normed assessments (Hill et al. 2008). Using this benchmark, pay-for-performance increased reading achievement in treatment schools by 8 percent of an average year of learning for students nationwide—equivalent to about 3 weeks of additional learning in a typical 36-week school year.

To facilitate comparisons between impacts on students' test scores in this analysis and the impacts on districts' measures of school achievement growth reported earlier, we used a method for approximately converting an impact on school achievement growth ratings into an implied impact on students' test scores (see Appendix B). Based on this conversion, the sizes of the impacts on student achievement in this analysis were similar to the sizes of the impacts on districts' measures of school achievement growth. Without regard to statistical significance, the impacts on student achievement found in this section were equivalent to one to three weeks of additional learning after Year 1 and about three weeks of additional learning after Year 2, depending on the subject examined. In comparison, the impacts on districts' measures of school achievement growth, when translated into implied impacts on student achievement, were equivalent to about three weeks of additional learning after both years.[70,71]

Pay-for-performance could affect elementary and middle school grades differently, so we examined impacts separately by grade span. For elementary and middle school students, impacts on reading achievement after Year 2 were nearly identical to the overall impact but were not statistically significant (Appendix F, Table F.17). Across all grade spans, subjects, and years, the only statistically significant impact of pay-for-performance was a positive impact on the reading achievement of middle school students after Year 1.

[69] This approximation is based on a normal distribution for student achievement.

[70] The impact on districts' measures of school achievement growth was roughly equivalent to raising student test scores by 0.04 student-level standard deviations after Year 1 and 0.03 student-level standard deviations after Year 2.

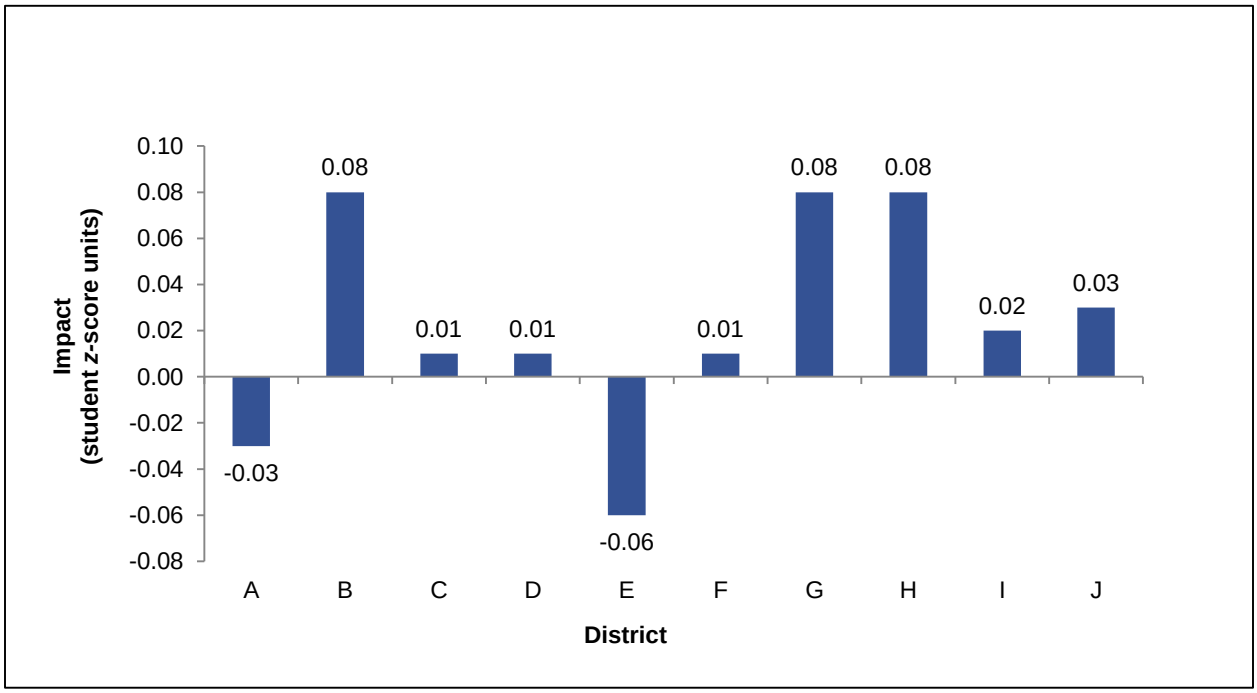
[71] The estimated impacts of pay-for-performance on student achievement were mostly consistent across a variety of alternative analytic models (see Appendix F, Tables F.12 through F.15). Some models that did not account for preexisting differences between treatment and control schools produced different findings. As discussed in Chapter II and Appendix B, our main analysis adjusted the impact findings to account for the fact that treatment schools had slightly lower student math achievement and slightly different student racial/ethnic composition than control schools at the beginning of the study. Failure to account for these preexisting differences could generate an inaccurate estimate of the effects of pay-for-performance. As expected, when we did not account for these preexisting differences, the estimated impacts of pay-for-performance on reading achievement were smaller and not statistically significant. In addition, after Year 1, neither the estimated impact on math nor the estimated impact on reading was statistically significant when Cohort 2 schools were included in the analysis (Appendix F, Table F.16).

## Differences in Student Achievement Impacts Across Districts

The findings shown in Table VI.3 represent an average impact of pay-for-performance across the 10 districts in the study. However, these districts differed in many ways, including the design and implementation of their pay-for-performance programs. These differences raise the possibility that the impacts of pay-for-performance could have also differed among districts. The data confirmed this was true.

**The impacts of pay-for-performance on math and reading achievement differed substantially across districts.** Although, on average, pay-for-performance had a positive impact on reading achievement, impacts varied across districts by a statistically significant degree (Figure VI.5). District-specific impacts on reading achievement after Year 2 ranged from -0.06 to 0.08 standard deviations and, without considering their statistical significance, impacts were positive in 8 of the 10 districts and negative in the other two. Impacts on math achievement after Year 2 also varied across districts, ranging from -0.11 to 0.25 standard deviations (Figure VI.6). Without considering their statistical significance, impacts in math were positive in half of the districts and negative in the other half.[72]

**Figure VI.5. Impact of Pay-for-Performance on Student Achievement in Reading After Year 2, by District (Student *z*-score units)**



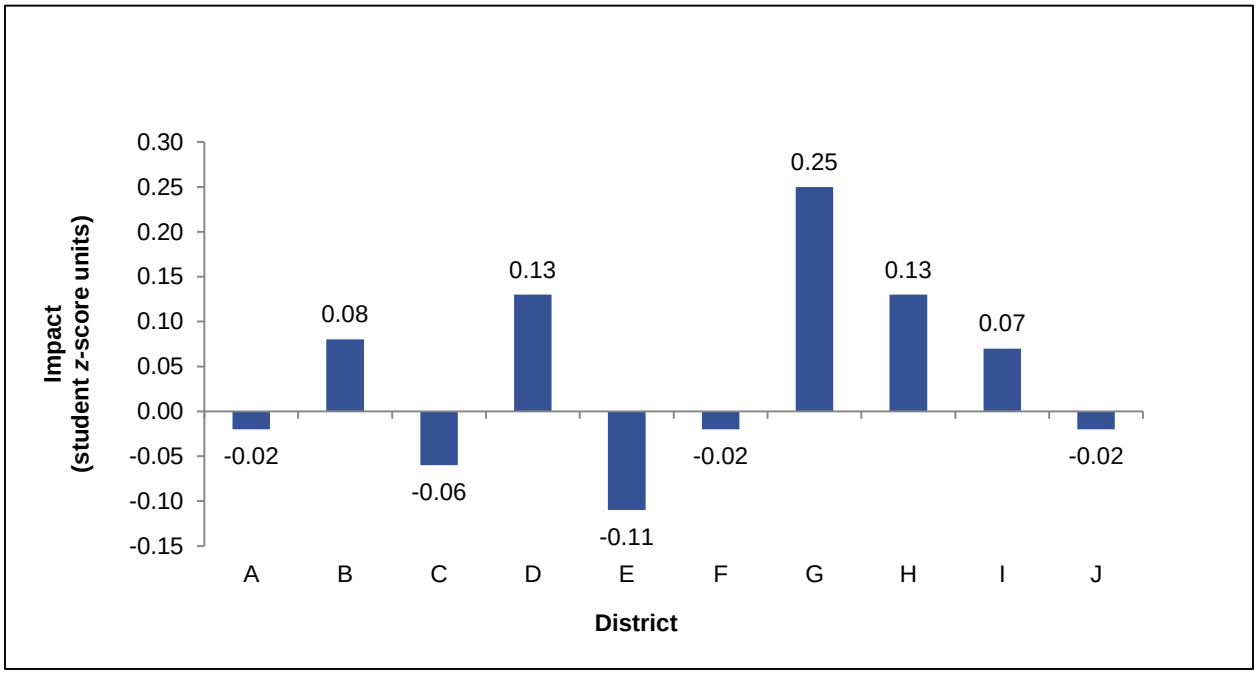

Source: Student administrative data (N = 40,391).

Note: An F-test of the null hypothesis that differences between treatment and control schools are equal across districts has a *p*-value of less than 0.001.

Figure reads: In District A, pay-for-performance lowered student reading achievement by 0.03 student *z*-score units after Year 2. In District B, pay-for-performance raised student reading achievement by 0.08 student *z*-score units after Year 2.

[72] Within each district, the small number of schools meant that only very large impacts would have been statistically significant. Therefore, we do not report the statistical significance of district-specific impacts, and instead focus on the overall variation in impacts across all 10 districts.

**Figure VI.6. Impact of Pay-for-Performance on Student Achievement in Math After Year 2, by District (Student *z*-score units)**




Source: Student administrative data (N = 40,709).

Note: An F-test of the null hypothesis that differences between treatment and control schools are equal across districts has a *p*-value of less than 0.001.

Figure reads: In District A, pay-for-performance lowered student math achievement by 0.02 student *z*-score units after Year 2.

We sought to identify explanations for why impacts differed across districts. In particular, as discussed in Chapter IV, both the design and implementation of TIF programs also differed across districts. Therefore, we examined whether impacts were systematically larger or smaller in districts that designed or implemented their programs in particular ways. Appendix G provides details on the methods and findings from this analysis.

**There was little evidence that key TIF program or implementation characteristics explain differences across districts in the impacts of pay-for-performance on student achievement.** The impacts of pay-for-performance on reading and math achievement were not related to a variety of program and implementation characteristics, including (1) the use of student achievement growth in teachers' own classrooms to measure teacher effectiveness and award bonuses, (2) teachers' understanding of their eligibility for performance bonuses, and (3) the timing of bonus notification and award (Appendix G, Table G.2). Only one program characteristic that we examined—the degree of differentiation in performance bonuses—had a statistically significant relationship with impacts. Higher differentiation in bonuses had a negative association with impacts on math achievement and no association with impacts on reading achievement.

## Summary

A primary objective of TIF grants is to raise student achievement in high-need schools. The evidence in this chapter indicates that the pay-for-performance component of TIF made a small contribution toward achieving this objective after the first and second years of TIF implementation. Pay-for-performance generated slightly higher student achievement in reading, with gains equivalent to about three weeks of additional learning. Gains in math were similar but not statistically significant.

The driving principle behind TIF is that increasing educator effectiveness is the key to raising student achievement and pay-for-performance bonuses are one way to increase educator effectiveness. Therefore, we examined whether the positive impact of pay-for-performance on student achievement was also reflected in positive impacts on educator effectiveness. On one measure of effectiveness that districts used to evaluate both teachers and principals—achievement growth in the educators' schools—we found consistent evidence that educators earned higher ratings as a result of pay-for-performance. This finding was not surprising, given that districts calculated school achievement growth ratings from some of the same student test scores for which we found positive impacts of pay-for-performance. However, despite the fact that pay-for-performance raised student reading achievement, it did not result in teachers or principals earning higher ratings on observations of their classroom or school practices. Observation ratings could have captured aspects of teaching or leadership quality that differed from those relevant for raising student achievement on state assessments. Moreover, as we documented in Chapter IV, the vast majority of teachers earned observation scores in the top half of the rating scale. Therefore, even improvements in teachers' classroom practices might not have led to a change in their observation rating.

Increases in educator effectiveness could have occurred either because teachers and principals improved their own effectiveness or because staffing changes resulted in more effective educators choosing to work at schools with pay-for-performance. We found little evidence for changes in staffing among teachers. Pay-for-performance did not enable schools to attract or retain more effective teachers. Among principals, we found some evidence that pay-for-performance caused more high performers to stay in their schools and more low performers to leave their schools after the first year of TIF implementation. However, it is unclear whether these staffing changes among principals actually contributed to the positive impacts of pay-for-performance on student achievement. The positive impacts on reading achievement materialized in the first year—before principals had the opportunity to remain in or leave their schools—and the impacts did not increase from the first to the second year. If schools that offer pay-for-performance continue to experience better retention of higher-performing principals, student achievement in these schools might increase further in the future. Before any of those changes are realized, the remaining explanation for why pay-for-performance raised student achievement in the first two years of TIF implementation is that it caused educators to improve their performance.

Given that the impacts of pay-for-performance on student achievement were small, one key question is whether any particular changes in how TIF programs were designed or implemented could enhance those impacts. In fact, we saw clear evidence that some districts realized larger impacts of pay-for-performance than others. However, these differences in impacts were not related to differences in several program and implementation characteristics that we measured. Therefore, we do not yet know what aspects of TIF programs are important for boosting the impacts of pay-for-performance on student achievement.

In the theory of change from Chapter I, educators' understanding of their TIF programs was thought to be essential for pay-for-performance to bring about improvements in educators' effectiveness. However, we found that the positive impact of pay-for-performance on reading achievement neither grew nor diminished from the first to the second year of TIF implementation, even though educators' understanding of how they were evaluated and whether they were eligible for performance bonuses improved (see Chapter IV). This suggests that educators' understanding of these program components was either not extensive enough or might not have been the critical factor determining the size of the impact on student achievement in the first two years of implementation. Evidence from future years will provide more clarity on whether, over a longer time period, the impacts of pay-for-performance evolve as educators continue accumulating more understanding of and experience with this program.

## REFERENCES


Angrist, Joshua D., and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2009.

Balch, Ryan, and Matthew Springer. "Performance Pay, Test Scores, and Student Learning Objectives." *Economics of Education Review*, vol. 44, 2015, pp. 114-125.

Bayonas, Holli. "Guilford County Schools Mission Possible Program: Year 3 (2008–09) External Evaluation Report." Greensboro, NC: University of North Carolina at Greensboro, SERVE Center, 2010.

Béteille, Tara, Demetra Kalogrides, and Susanna Loeb. "Stepping Stones: Principal Career Paths and School Outcomes." *Social Science Research,* vol. 41, no. 4, 2012, pp. 904–919.

Dee, Thomas, and James Wyckoff. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." Working paper no. 19529. Cambridge, MA: National Bureau of Economic Research, October 2013.

Donald, Stephen G., and Kevin Lang. "Inference with Difference-in-Differences and Other Panel Data." *Review of Economics and Statistics*, vol. 89, 2007, pp. 221–233.

Freedman, David A. "On Regression Adjustments to Experimental Data." *Advances in Applied Mathematics*, vol. 40, 2008, pp. 180–193.

Fryer, Roland. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." Working paper no. 16850. Cambridge, MA: National Bureau of Economic Research, March 2011.

Fryer, Roland, Steven Levitt, John List, and Sally Sadoff. "Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment." Working paper no. 18237. Cambridge, MA: National Bureau of Economic Research, July 2012.

Fulbeck, Eleanor. "Teacher Mobility and Financial Incentives: A Descriptive Analysis of Denver's ProComp." *Educational Evaluation and Policy Analysis*, vol. 36, no. 1, March 2014, pp. 67–82.

Gates, Susan M., Jeanne S. Ringel, Lucrecia Santibanez, Cassandra Guarino, Bonnie Ghosh-Dastidar, and Abigail Brown. "Mobility and Turnover Among School Principals." *Economics of Education Review*, vol. 25, no. 3, June 2006, pp. 289–302.

Glazerman, Steven, Allison McKie, and Nancy Carey. "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year One Impact Report." Washington, DC: Mathematica Policy Research, April 2009.

Glazerman, Steven, and Allison Seifullah. "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact Report." Washington, DC: Mathematica Policy Research, May 2010.

Glazerman, Steven, Hanley Chiang, Alison Wellington, Jill Constantine, and Daniel Player. "Impacts of Performance Pay Under the Teacher Incentive Fund: Study Design Report." Final report submitted to the U.S. Department of Education, Institute of Education Sciences. Washington, DC: Mathematica Policy Research, October 2011.

Glazerman, Steven, and Allison Seifullah. "An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) After Four Years." Washington, DC: Mathematica Policy Research, March 2012.

Goldhaber, Dan, and Joe Walch. "Strategic Pay Reform: A Student Outcomes-Based Evaluation of Denver's ProComp Teacher Pay Initiative." *Economics of Education Review,* vol. 31, 2012, pp. 1067–1083.

Goodman, Sarena, and Lesley Turner. "Does Whole School Performance Pay Improve Student Learning? Evidence from the New York City Schools." *Education Next*, vol. 11, no. 2, spring 2011.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*, vol. 2, no. 3, 2008, pp. 172–177.

Kamenica, Emir. "Behavioral Economics and Psychology of Incentives." *Annual Review of Economics,* vol. 4, 2012, pp. 427–452.

Keigher, Ashley. "Teacher Attrition and Mobility: Results from the 2008-09 Teacher Follow-up Survey (NCES 2010-353)." Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2010.

Liang, Kung-Yee, and Scott L. Zeger. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, vol. 73, 1986, pp. 13–22.

Marsh, J.A., M.G. Springer, D.F. McCaffrey, K. Yuan, S. Epstein, J. Koppich, and A. Peng. *A Big Apple for Educators: New York City's Experiment with Schoolwide Performance Bonuses.* Final evaluation report. Santa Monica, CA: RAND Corporation, 2011.

Max, Jeffrey, Jill Constantine, Alison Wellington, Kristin Hallgren, Steven Glazerman, Hanley Chiang, and Cecilia Speroni. "Evaluation of the Teacher Incentive Fund: Implementation and Early Impacts of Pay-for-Performance After One Year." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, September 2014.

Puma, Michael, Robert Olsen, Stephen Bell, and Cristofer Price. "What to Do When Data Are Missing in Group Randomized Controlled Trials." NCEE 2009-0049. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, October 2009.

Ringel, Jeanne, Susan Gates, Catherine Chung, Abigail Brown, and Bonnie Ghosh-Dastidar. "Career Paths of School Administrators in Illinois: Insights from an Analysis of State Data." Santa Monica, CA: RAND Corporation, 2004.

Rubin, Donald. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons, 1987.

Schafer, Joseph, and John Graham. "Missing Data: Our View of the State of the Art." *Psychological Methods*, vol. 7, 2002, pp. 147–177.

Schenker, Nathaniel, and Jeremy Taylor. "Partially Parametric Techniques for Multiple Imputation." *Computation Statistics & Data Analysis*, vol. 22, 1996, pp. 425–446.

Shifrer, Dara, Ruth Turley, and Holly Heard. "Houston Independent School District's ASPIRE Program: Estimated Effects of Receiving Financial Awards." Working paper. Houston, TX: Houston Education Research Consortium, 2013.

Slotnik, William, Maribeth Smith, Barbara Helms, and Zhaogang Qiao. "It's More than Money: Teacher Incentive Fund—Leadership for Educators' Advanced Performance. Charlotte-Mecklenburg Schools." Boston, MA: Community Training and Assistance Center, February 2013.

Sojourner, Aaron, Elton Mykerezi, and Kristine West. "Teacher Pay Reform and Productivity: Panel Data Evidence from Adoptions of Q-Comp in Minnesota." *Journal of Human Resources*, vol. 49, no. 4, 2014, pp. 945–981.

Springer, Matthew, Jessica Lewis, Michael Podgursky, Mark Ehlert, Lori Taylor, Omar Lopez, and Art (Xiao) Peng. "Governor's Educator Excellence Grant (GEEG) Program: Year Three Evaluation Report." Nashville, TN: National Center on Performance Incentives, 2009a.

Springer, Matthew, Jessica Lewis, Michael Podgursky, Mark Ehlert, Timothy Grownberg, Laura Hamilton, Dennis Jansen, Brian Stecher, Lori Taylor, Omar Lopez, and Art (Xiao) Peng. "Texas Educator Excellence Grant (TEEG) Program: Year Three Evaluation Report." Nashville, TN: National Center on Performance Incentives, 2009b.

Springer, Matthew, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel McCaffrey, Matthew Pepper, and Brian Stecher. "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching." Nashville, TN: Vanderbilt University, National Center on Performance Incentives, 2010.

Springer, Matthew, John Pane, Vi-Nhuan Le, Daniel McCaffrey, Susan Burns, Laura Hamilton, and Brian Stecher. "Team Pay for Performance: Experimental Evidence from the Round Rock Pilot Project on Team Incentives. *Educational Evaluation and Policy Analysis,* vol. 34, no. 4, 2012, pp. 367–390.

Springer, Matthew, Dale Ballou, and Art Peng. "Estimated Effect of the Teacher Advancement Program on Student Test Score Gains: Findings from an Independent Appraisal." *Education Finance and Policy,* vol. 9, no. 2, 2014, pp. 193–230.

THIS PAGE IS INTENTIONALLY BLANK

# APPENDIX A

# SUPPLEMENTAL INFORMATION ON STUDY SAMPLE, DESIGN, DATA, AND METHODS FOR CHAPTER II

THIS PAGE IS INTENTIONALLY BLANK

This appendix provides more detailed information about characteristics of TIF districts, the study design, the teacher survey sample, survey response rates, and sample sizes for analyses using educator and student administrative data.

As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts whose schools were randomly assigned in spring and summer 2011 were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 1 completed two years of implementation during the period covered by this report, 2011–2012 and 2012–2013, referred to as Years 1 and 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort.

## Random Assignment of Schools to the Treatment and Control Groups

To randomly assign schools within a district to the treatment and control groups, we used a matched-pair randomization approach designed to maximize the balance between the treatment and control groups on observable characteristics. Specifically, we used two approaches: (1) creating matched pairs of schools, and (2) creating matched groups of schools.

**Matched pairs of schools.** We randomly assigned most of the schools (72 of 138 Cohort 1 schools, and 42 of 45 Cohort 2 schools) to treatment and control groups within matched pairs of schools. One school in each pair was randomly selected to be in the treatment group; the other school was assigned to the control group. Within each district, pairs were constructed so the schools that were paired together would (1) have identical sets of grades represented; (2) be similar in average student achievement; and (3) be similar on other characteristics, such as school size, percentage of students eligible for free or reduced-price lunch, and racial/ethnic composition. District staff either approved the pairs that we constructed or directly specified the pairs based on their knowledge of the participating schools. Because pairing reduced the chance that randomization would produce treatment and control groups with large baseline differences, it enhanced precision for estimating the impacts of pay-for-performance bonuses.

**Matched groups of schools.** For the remaining schools (66 of 138 Cohort 1 schools, and 3 of 45 Cohort 2 schools), we randomly assigned groups of schools to treatment and control groups within matched pairs of groups. This was analogous to the matched-pairs procedure described previously, except that we assigned groups of schools within matched pairs of groups rather than assigning individual schools within matched pairs of individual schools. We used this approach when the randomization had to satisfy constraints that could not be met with paired random assignment of individual schools. For example, some districts requested that certain schools be assigned to the same treatment status if they were expected to be consolidated in the future or were in the same feeder pattern (for instance, grouping a middle school with the elementary schools from which its students typically came). Moreover, in some districts, all participating schools in the district were grouped into two groups that were well matched on average baseline characteristics; this was done to address concerns that several individual schools would not have had suitable matches if pairs of individual schools had been constructed. As with the pairing of individual schools described earlier, the pairing of groups of schools was designed to minimize the chance that randomization would produce treatment and schools that were dissimilar on baseline characteristics.

## School Attrition

For our primary analysis in Chapters IV through VI, we focus on Cohort 1 schools that had implemented TIF for two full years (Year 1 is 2011–2012 and Year 2 is 2012–2013). Of the 138 Cohort 1 schools that were randomly assigned, 6 schools were dropped from all analyses to keep a constant analysis sample of 132 schools each year. After the first year of TIF implementation, four schools either closed, chose to drop out of the study, or were consolidated. These schools, along with their matched pair, are excluded from the main analysis. However, the results based on Cohorts 1 and 2 (shown in later appendices) include these schools in the analyses for Year 1.

As Table A.1 shows, school attrition was low, ranging from 4.3 to 5.1 percent for Cohort 1 and 0 to 0.5 percent for Cohorts 1 and 2. There was no difference in the attrition rate between treatment and control schools.

**Table A.1. School Attrition, Cohorts 1 and 2 (Percentages Unless Otherwise Noted)**

| | Overall | Treatment | Control | Differential Attrition |
|---|---|---|---|---|
| **Cohort 1** | | | | |
| **Number of Schools Randomly Assigned** | 138 | 69 | 69 | NA |
| Analyses of Student and Educator Administrative Data | | | | |
| Number of Schools in Year 1 Analyses[a] | 132 | 66 | 66 | NA |
| Number of Schools in Year 2 Analyses[b] | 132 | 66 | 66 | NA |
| Number of Schools in Year 3 Analyses[c] | 132 | 66 | 66 | NA |
| Attrition Rate Year 1 | 4.3 | 4.3 | 4.3 | 0 |
| Attrition Rate Year 2 | 4.3 | 4.3 | 4.3 | 0 |
| Attrition Rate Year 3[a] | 4.3 | 4.3 | 4.3 | 0 |
| Analyses of Educator Survey Data | | | | |
| Number of Schools in Year 1 Analyses | 131 | 66 | 65 | NA |
| Number of Schools in Year 2 Analyses | 132 | 66 | 66 | NA |
| Attrition Rate Year 1 | 5.1 | 4.3 | 5.6 | -1.3 |
| Attrition Rate Year 2 | 4.3 | 4.3 | 4.3 | 0 |
| **Cohorts 1 and 2** | | | | |
| **Number of Schools Randomly Assigned** | 183 | 92 | 91 | NA |
| Analyses of Student and Educator Administrative Data | | | | |
| Number of Schools in Year 1 Analyses[a] | 183 | 92 | 91 | NA |
| Attrition Rate Year 1 | 0 | 0 | 0 | 0 |
| Analyses of Educator Survey Data | | | | |
| Number of Schools in Year 1 Analyses | 182 | 92 | 90 | NA |
| Attrition Rate Year 1 | 0.5 | 0 | 1.1 | -1.1 |

Notes: The primary analyses in the main body of the report are based on schools that implemented the program for two years (Cohort 1). Supplemental analyses are based on all study schools that implemented the program for at least one year (Cohorts 1 and 2) and are reported in the appendices.

[a]Includes analyses of student achievement and educator performance ratings in Year 1.

[b]Includes analyses of student achievement and educator performance ratings in Year 2, and educator retention from Year 1 to Year 2.

[c]Includes analyses of educator retention from Year 1 to Year 3.

NA is not applicable.

*Differential attrition is statistically significant at the .05 level, two-tailed test.

## Baseline Characteristics of Treatment and Control Schools

By virtue of random assignment, treatment and control schools should have similar characteristics at the time of randomization. In Chapter II, we examined whether random assignment produced treatment and control groups that were equivalent at the beginning of the study (the 2010–2011 school year) for the Cohort 1 schools in our main analyses. Tables A.2 and A.3 show similar information for study schools in Cohorts 1 and 2. The samples sizes in these tables are smaller than the full sample sizes due to missing data. For example, districts did not provide data on educator or student characteristics for some schools in our study, so school sample sizes in these tables are smaller than the full sample of Cohort 1 and 2 schools (183 schools).

**Table A.2. Average Baseline Characteristics of Students Enrolled in Treatment and Control Schools in the Pre-Implementation Year, Cohorts 1 and 2 (Percentages Unless Otherwise Noted)**

| | Treatment | Control | Difference |
|---|---|---|---|
| Achievement in the Pre-Implementation Year (*z*-score) | | | |
| Math | -0.57 | -0.52 | -0.05* |
| Reading | -0.52 | -0.49 | -0.03 |
| Race/Ethnicity | | | |
| White, non-Hispanic | 24 | 27 | -3* |
| African American, non-Hispanic | 47 | 46 | 1 |
| Hispanic | 22 | 21 | 2* |
| Other | 6 | 6 | 0 |
| Other Characteristics | | | |
| Female | 49 | 49 | -1 |
| Eligible for free/reduced-price lunch | 81 | 79 | 1 |
| Disabled or has an Individualized Education Program | 14 | 14 | 0 |
| Overage for grade | 14 | 14 | 0 |
| English language learner | 9 | 8 | 0 |
| Grade Span | | | |
| Grades 3–5 | 61 | 61 | 0 |
| Grades 6–8 | 39 | 39 | 0 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.049 |
| **Number of Students—Range[a]** | **20,033-31,083** | **19,592-30,482** | |
| **Number of Schools—Range[a]** | **63-91** | **62-90** | |

Source: Student administrative data.

Notes: The table is based on the 181 Cohort 1 and Cohort 2 study schools. The pre-implementation year is 2010–2011 for Cohort 1 and 2012–2013 for Cohort 2. One school did not provide data for the pre-implementation year, so we excluded this school and its matched school from this analysis. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.3. Average Characteristics of Educators in Treatment and Control Schools in Year 1, Cohorts 1 and 2 (Percentages Unless Otherwise Noted)**

| | Teachers | | | Principals | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 83 | 83 | 0 | 65 | 60 | 5 |
| Race/ethnicity | | | | | | |
| White, non-Hispanic | 75 | 75 | 0 | 59 | 55 | 4 |
| Black, non-Hispanic | 18 | 19 | -1 | 34 | 38 | -4 |
| Hispanic | 3 | 3 | 0 | 4 | 2 | 1 |
| Other | 4 | 4 | 0 | 4 | 5 | -1 |
| Age (average years) | 42 | 42 | 1* | 49 | 48 | 2 |
| Education | | | | | | |
| Master's degree or higher | 59 | 58 | 1 | 95 | 95 | 1 |
| Experience in K–12 education | | | | | | |
| Total experience (average years) | 12 | 12 | 0 | 16 | 14 | 2 |
| Less than 5 years | 24 | 24 | 0 | 20 | 14 | 6 |
| 5-15 years | 44 | 46 | -1 | 31 | 43 | -12 |
| More than 15 years | 32 | 31 | 2 | 49 | 43 | 6 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.911 | | | 0.539 |
| **Number of Educators—Range[a]** | **2,268-3,027** | **2,260-2,888** | | **53-84** | **57-88** | |
| **Number of Schools—Range[a]** | **72-91** | **71-90** | | **50-81** | **55-84** | |

Source: Educator administrative data.

Notes: Year 1 is 2011–2012 for Cohort 1 and 2012–2013 for Cohort 2. The number of principals exceeds the number of schools in the analysis sample because a few schools had more than one principal. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

We lacked baseline data on educators for one of the 10 Cohort 1 districts; therefore, in Chapter II, we showed educator characteristics at the beginning of Year 1. Of the 132 Cohort 1 schools in the final analysis sample, 20 were in the district that did not provide pre-implementation information. Table A.4 shows pre-implementation characteristics for the 112 schools in the nine Cohort 1 districts that provided us with educator characteristics in the pre-implementation year.

**Table A.4. Average Characteristics of Educators in Treatment and Control Schools in the Pre-Implementation Year, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Teachers | | | Principals | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 86 | 85 | 1 | 66 | 56 | 9 |
| Race/ethnicity | | | | | | |
| White, non-Hispanic | 76 | 73 | 3* | 70 | 63 | 7 |
| Black, non-Hispanic | 17 | 20 | -3* | 24 | 30 | -6 |
| Hispanic | 2 | 2 | 0 | 2 | 2 | 0 |
| Other | 5 | 5 | 0 | 4 | 5 | -1 |
| Age (average years) | 43 | 43 | 0 | 48 | 48 | 0 |
| Education | | | | | | |
| Master's degree or higher | 58 | 59 | -2 | 99 | 90 | 8 |
| Experience in K–12 Education | | | | | | |
| Total experience (average years) | 13 | 13 | 0 | 16 | 15 | 1 |
| Less than 5 years | 20 | 19 | 1 | 11 | 10 | 1 |
| 5-15 years | 46 | 47 | 0 | 43 | 43 | 0 |
| More than 15 years | 34 | 34 | 0 | 46 | 47 | -1 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.036 | | | 0.000 |
| **Number of Educators—Range[a]** | **729-1,732** | **770-1,722** | | **25-50** | **28-53** | |
| **Number of Schools—Range[a]** | **27-56** | **27-56** | | **24-49** | **26-51** | |

Source: Educator administrative data.

Notes: One district did not provide data for the pre-implementation year. The number of principals exceeds the number of schools in the analysis sample because a few schools had more than one principal. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

## Selection of the Teacher Survey Sample

As discussed in Chapter II, we surveyed a subset of the teachers in all of the study schools that were randomized in spring and summer 2011 (Cohort 1 schools) or in spring and summer 2012 (Cohort 2 schools). Here, we describe the rationale for the specific grades and subjects included in our sample and our methods for selecting the teachers to whom we administered the 2012 and 2013 teacher surveys.

### Teaching Assignments Targeted by the Surveys

For the teacher surveys, we targeted teachers who taught 1st grade, 4th grade, 7th-grade math, 7th-grade English/language arts, or 7th-grade science in the study schools. We decided to focus on specific grades and subjects, rather than all elementary and middle school grades and subjects, to minimize the chance that the grades and subjects represented in the teacher sample would differ substantially between the treatment and control schools that were compared in the analysis. In other words, we wanted any treatment-control differences in teacher-reported outcomes to be attributable to pay-for-performance, rather than to an imbalance in grades or subjects.

We chose these grades and subjects so that they would encompass different groups of teachers who were thought to face different incentives from pay-for-performance—in particular, teachers in tested grade/subject combinations (4th grade, 7th-grade math, and 7th-grade reading)—and those in nontested grade/subject combinations (1st grade and 7th-grade science). Teachers in nontested grades/subjects might be eligible for bonuses based heavily on performance measures that they could affect only indirectly (such as student achievement growth in other grades and subjects within the same school). On the other hand, teachers in tested grades/subjects could have a more direct influence on performance ratings—and, therefore, bonus amounts—that were linked to the achievement growth of students in their own classrooms.

The set of targeted grades was also designed to include both elementary and middle school grades because of their different classroom structures. Elementary school teachers typically teach self-contained classrooms and are responsible for all core subjects, whereas middle school teachers typically work in a departmentalized setting in which they are responsible for one subject (such as math *or* reading). Among the tested elementary grades, we chose to target 4th grade because it is typically the earliest grade at which student achievement growth on state assessments can be calculated and is more likely than grade 5 to have self-contained classes. Among the tested middle school grades and subjects, we chose 7th-grade math and reading because they are more likely than 8th-grade subjects to be assessed by end-of-grade tests that are uniform across all students (rather than end-of-course tests that depend on the course in which students are enrolled) but are more likely than 6th-grade classes to be departmentalized.

We chose 1st grade and 7th-grade science as the nontested grades and subjects in our target population, for several reasons. First grade has full-day classes and is less likely than grades 2 and 3 to have standardized testing. Science is a well-defined subject that is not tested annually, and retaining certified science teachers is an important policy goal.

**Sampling Approach**

Although both 2012 and 2013 surveys focused on teachers in the targeted grades and subjects described above, there were some differences in the sampling approach used each year. Specifically, in 2013 we sampled (1) all teachers in targeted grades and subjects (as opposed to a subset of them), and (2) teachers who were surveyed in 2012 even if they were no longer teaching a targeted grade and subject.

**Sampling approach for teachers in targeted grades and subjects.** Within each study school and year, we used administrative data provided by the evaluation districts to identify teachers who were assigned to any of the targeted grades and subjects. In 2012, we sampled all 4th-grade teachers; all 7th-grade math, English/language arts, and science teachers; and 77 percent of 1st-grade teachers. Because our analysis of impacts on student achievement focuses on tested grades and subjects, our sampling approach for the teacher survey was designed to give greater emphasis to tested grades and subjects than to nontested ones. Therefore, we selected all teachers who taught any of the tested grades and subjects targeted by the survey and selected a subset of teachers who taught the nontested grades and subjects targeted by the survey. Specifically, for each nontested grade and subject (1st grade or 7th-grade science) in each study school, we randomly selected three teachers from the teachers assigned to that combination of school, grade, and subject. If no more than three teachers were assigned to that combination, all such teachers were chosen. In practice, this approach led to the selection of all 7th-grade science teachers in the sampling frame—due to the small numbers of such

teachers in each school—and 77 percent of the 1st-grade teachers in the sampling frame.[73] In 2013, we surveyed all teachers in targeted grades and subjects, including 100 percent of 1st-grade teachers, which led to an increase in the total number of teachers in these targeted teaching assignments.

**Sampling approach for teachers previously surveyed.** In 2013, we also sampled those teachers who were surveyed in 2012 but were no longer teaching a targeted grade and subject. If pay-for-performance had an impact on teachers' school choice or career decisions, this subset of teachers would have allowed us to document reasons why teachers switch schools or leave the teaching profession.

We wanted to survey teachers from two groups of teachers: (1) teachers in the targeted grades and subjects, and (2) teachers we had surveyed the year before but were no longer teaching a targeted grade or subject. However, because some teaching rosters were not sufficiently detailed (for example, describing teachers' grades as a range of grades) or were inaccurate, our sample included 97 teachers in 2012 and 113 in 2013 who reported they were not teaching in the targeted grades and subjects, although we had believed they were. We excluded these teachers from the Year 2 teacher survey analyses. We did not need to replace these ineligible teachers because we had already selected all teachers identified by the administrative data as teaching the grades and subjects targeted by the survey. Similarly, some teachers we surveyed in 2013 because we had surveyed them in 2012, although we thought they no longer were teaching in a targeted grade and subject, reported they were teaching a targeted grade and subject. We included these teachers' responses in our Year 2 teacher survey analyses.

## Survey Response Rates and Analysis of Missing Outcomes in Survey Data

In this section, we report the response rates for each of the three surveys (district, teacher, and principal surveys) and years used in this report. Because of the high response rate (more than 88 percent across all surveys), the potential for nonresponse bias is minimal. Nonetheless, we assessed the extent to which the respondents are similar to nonrespondents and, for educator surveys, whether respondents are similar across treatment and control schools.

Table A.5 shows the response rates for the 2013 district survey, and Table A.6 compares district characteristics of respondents and nonrespondents on such dimensions as district location and size.

**Table A.5. District Survey Response Rates Overall and by Evaluation Status, 2012–2013 School Year**

| | Overall | Non-Evaluation Districts | Evaluation Districts |
|---|---|---|---|
| All Districts | | | |
| Number of districts | 169 | 156 | 13 |
| Number of respondents | 160 | 147 | 13 |
| Number of ineligible respondents | 5 | 5 | 0 |
| Response rate (respondents over total) | 95 | 94 | 100 |

Source: District survey, 2013.

Notes: Ineligible districts are districts that indicated in the survey they were not implementing TIF at the time of survey administration. The difference in response rates between non-evaluation and evaluation districts was not statistically significant at the .05 level.

[73] Due to an error in the sampling algorithm, we inadvertently sampled all 1st-grade teachers in three districts' study schools.

**Table A.6. District Characteristics by Districts' Response Status, 2012–2013 School Year (Percentages Unless Otherwise Noted)**

| | Respondents | Nonrespondents |
|---|---|---|
| Student Racial/Ethnic Distribution | | |
| White, non-Hispanic | 49 | 20 |
| Black, non-Hispanic | 25 | 52 |
| Hispanic | 20 | 23 |
| Student Socioeconomic Status | | |
| Eligible for free/reduced-price lunch | 63 | 60 |
| Title 1 eligible schools (schoolwide) | 76 | 76 |
| Enrollment (averages) | | |
| Total enrollment | 20,739 | 16,068 |
| District Location[a] | | |
| Urban | 36 | 29* |
| Suburban | 12 | 57 |
| Town | 20 | 0 |
| Rural | 32 | 14* |
| District Census Bureau Region | | |
| Northeast | 9 | 0 |
| Midwest | 28 | 0 |
| South | 43 | 89 |
| West | 20 | 11* |
| **Number of Districts** | **150-160** | **6-9** |

Source: District survey (2013) and Common Core of Data for 2011–2012 school year.

Notes: Ten TIF non-evaluation districts are not included in the 2011–2012 district-level data from the Common Core of Data. Common Core of Data school-level data are used to calculate socioeconomic indicators. Common Core of Data district-level data are used to calculate all other demographic characteristics.

[a]District location indicates the physical location of the district agency.

*Difference between respondents and nonrespondents is statistically significant at the .05 level, two-tailed test.

Tables A.7 and A.8 show teacher and principal sample sizes and response rates. Table A.7 reports the total number of surveyed teachers in 1st grade, 4th grade, and 7th-grade math, English/language arts, and science and principals in Cohort 1 schools, along with their response rates and the final analyses samples. Table A.8 shows response rates for teachers (those in targeted grades and subjects) in Cohort 2.

**Table A.7. Teacher and Principal Response Rates for the Final Analyses Samples, Cohort 1**

| | Year 1 (2012 Survey) | | | Year 2 (2013 Survey) | | |
|---|---|---|---|---|---|---|
| | Total | Treatment | Control | Total | Treatment | Control |
| **Teachers** | | | | | | |
| Number of Sampled Teachers[a] | 961 | 478 | 483 | 950 | 471 | 479 |
| Number of respondents | 880 | 433 | 447 | 872 | 433 | 439 |
| Response rate (percentage) | 92 | 91 | 93 | 92 | 92 | 92 |
| Number of teachers in the final analysis sample[b] | 795 | 393 | 402 | 904 | 451 | 453 |
| **Principals** | | | | | | |
| Number of Sampled Principals | 132 | 66 | 66 | 132 | 66 | 66 |
| Number of respondents | 129 | 65 | 64 | 126 | 64 | 62 |
| Response rate (percentage) | 98 | 98 | 97 | 95 | 97 | 94 |
| Number of principals in the final analysis sample[c] | 129 | 65 | 64 | 125 | 64 | 61 |

Source: Teacher and principal surveys, 2012 and 2013.

Note: None of the differences in response rates between educators in treatment and control schools were statistically significant at the .05 level.

[a]The teacher sample for the final analysis included 1st grade, 4th grade, and 7th-grade math, English/language arts, and science teachers.

[b]The final analysis sample excludes teachers who reported working part-time or teaching grades and subjects other than the targeted 1st grade, 4th grade, and 7th-grade math, English/language arts, and science. In addition, it includes teachers who were not in our original sample of teachers in targeted grades and subjects but who responded to the survey and self-identified as teaching in those targeted grades and subjects.

[c]The analysis sample in Year 2 excludes a few respondents who did not identify themselves as principals in the survey.

**Table A.8. Teacher and Principal Response Rates for the Final Analyses Samples, Cohort 2**

| | Year 1 (2013 Survey) | | |
|---|---|---|---|
| | Total | Treatment | Control |
| **Teachers** | | | |
| Number of Sampled Teachers[a] | 259 | 139 | 120 |
| Number of respondents | 237 | 124 | 113 |
| Response rate (percentage) | 92 | 89 | 94 |
| Number of teachers in the final analysis sample[b] | 251 | 136 | 115 |
| **Principals** | | | |
| Number of Sampled Principals | 45 | 23 | 22 |
| Number of respondents | 39 | 20 | 19 |
| Response rate (percentage) | 87 | 87 | 86 |
| Number of teachers in the final analysis sample[c] | 38 | 19 | 19 |

Source: Teacher and principal surveys, 2013.

Note: None of the differences in response rates between educators in treatment and control schools were statistically significant at the .05 level.

[a]The teacher sample for the final analysis included 1st grade, 4th grade, and 7th-grade math, English/language arts, and science teachers.

[b]The final analysis sample excludes teachers who reported working part-time or teaching grades and subjects other than the targeted 1st grade, 4th grade, and 7th-grade math, English/language arts, and science. In addition, it includes teachers who were not in our original sample of teachers in targeted grades and subjects but who responded to the survey and self-identified as teaching in those targeted grades and subjects.

[c]The analysis sample excludes a few respondents who did not identify themselves as principals in the survey.

Table A.9 presents the distribution of grade and subject assignments for the Cohort 1 teachers who responded to the survey and were included in the final analysis samples.

**Table A.9. Teacher Respondents, by Teaching Assignment and Treatment Status, Cohort 1**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Grade Taught | Total | Treatment | Control | Total | Treatment | Control |
| 1st grade only | 226 | 109 | 117 | 302 | 157 | 145 |
| 4th grade only | 222 | 111 | 111 | 220 | 105 | 115 |
| 7th-grade English/language arts and/or math only | 203 | 100 | 103 | 199 | 98 | 101 |
| 7th-grade science only | 66 | 37 | 29 | 60 | 34 | 26 |
| More than one targeted grade or subject | 78 | 36 | 42 | 123 | 57 | 66 |
| **Total** | **795** | **393** | **402** | **904** | **451** | **453** |

Source: Teacher survey, 2012 and 2013.

Notes: Targeted grades and subjects for the survey were 1st grade, 4th grade, and 7th-grade math, English/language arts, and science. Counts are for teachers in those targeted grades and subjects who responded to the survey.

We matched administrative data to survey respondents to compare (1) the characteristics of respondents and nonrespondents, and (2) the characteristics of educators in treatment and control schools. Tables A.10 through A.12 present our nonresponse analyses for the teacher and principal surveys. Table A.10 compares the characteristics of teachers who responded to the survey to those who did not. Because there were few principal nonrespondents, we do not report a similar analysis for the principal survey. Tables A.11 and A.12 compare the characteristics of respondents in treatment and control schools for teachers and principals, respectively. Because we did not receive administrative data on educator characteristics for all survey respondents, the sample sizes in Tables A.10 through A.12 are smaller than the number of teacher and principal survey respondents.

**Table A.10. Characteristics of Teacher Survey Respondents and Nonrespondents, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| | Respondents | Nonrespondents | Respondents | Nonrespondents |
| Demographic Characteristics | | | | |
| Female | 89 | 78 | 87 | 84 |
| Race/Ethnicity | | | | |
| White, non-Hispanic | 71 | 66 | 71 | 72 |
| Black, non-Hispanic | 22 | 27 | 22 | 19 |
| Hispanic | 2 | 0 | 4 | 3 |
| Other | 4 | 7 | 3 | 6 |
| Age (average years) | 40 | 43* | 41 | 43 |
| Education | | | | |
| Master's degree or higher | 47 | 34 | 57 | 46 |
| Experience in K–12 Education | | | | |
| Total experience (average years) | 12 | 14 | 11 | 11 |
| Less than 5 years | 22 | 22 | 28 | 29 |
| 5-15 years | 44 | 35 | 45 | 45 |
| More than 15 years | 34 | 44* | 27 | 26 |
| **Number of Teachers—Range[a]** | **540-734** | **45-70** | **650-867** | **39-65** |

Source: Teacher surveys (2012 and 2013) and educator administrative data.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference between respondents and nonrespondents is statistically significant at the .05 level, two-tailed test.

**Table A.11. Characteristics of Teacher Survey Respondents by Treatment Status, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| | Treatment | Control | Treatment | Control |
| Demographic Characteristics | | | | |
| Female | 88 | 85 | 89 | 87 |
| Race/Ethnicity | | | | |
| White, non-Hispanic | 75 | 67* | 72 | 69 |
| Black, non-Hispanic | 18 | 24* | 21 | 25* |
| Hispanic | 3 | 3 | 4 | 2 |
| Other | 4 | 6 | 4 | 3 |
| Age (average years) | 40 | 40 | 41 | 41 |
| Education | | | | |
| Master's degree or higher | 44 | 52* | 48 | 55 |
| Experience in K–12 Education | | | | |
| Total Experience (average years) | 11 | 10 | 11 | 10 |
| Less than 5 years | 27 | 27 | 26 | 30 |
| 5-15 years | 46 | 49 | 47 | 46 |
| More than 15 years | 27 | 23 | 27 | 24 |
| **Number of Teachers—Range[a]** | **248-357** | **292-377** | **326-436** | **324-431** |

Source: Teacher surveys (2012 and 2013) and educator administrative data.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.12. Characteristics of Principal Survey Respondents by Treatment Status, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| | Treatment | Control | Treatment | Control |
| Demographic Characteristics | | | | |
| Female | 60 | 63 | 65 | 64 |
| Race/Ethnicity | | | | |
| White, non-Hispanic | 66 | 60 | 60 | 52 |
| Black, non-Hispanic | 26 | 32 | 34 | 36 |
| Hispanic | 2 | 0 | 2 | 4 |
| Other | 6 | 8 | 4 | 8 |
| Age (average years) | 50 | 48 | 48 | 48 |
| Education | | | | |
| Master's degree or higher | 95 | 92 | 100[b] | 94 |
| Experience in K–12 Education | | | | |
| Total experience (average years) | 17 | 16 | 16 | 13* |
| Less than 5 years | 13 | 10 | 17 | 20 |
| 5-15 years | 37 | 40 | 27 | 42 |
| More than 15 years | 50 | 50 | 56 | 38* |
| **Number of Principals—Range[a]** | **37-54** | **36-53** | **46-57** | **34-50** |

Source: Principal surveys (2012 and 2013) and educator administrative data.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

[b]Regression-adjusted value was above 100 and was therefore top-coded to be 100.

*Difference is statistically significant at the .05 level, two-tailed test.

## Sample Sizes and Analysis of Missing Outcomes in Educator Administrative Data

We used districts' administrative records for all analyses of educator effectiveness. In this section, we describe the samples and the characteristics of educators included in these analyses.

All analyses of educator effectiveness were restricted to educators who worked full-time in the study schools. The 132 Cohort 1 schools included 4,346 full-time teachers in Year 1 and 4,466 full-time teachers in Year 2. The number of full-time principals was not the same as the total number of study schools because a few schools did not have a full-time principal or had more than one full-time principal. Table A.13 shows the number of full-time principals listed in the administrative data and the number of schools in those principals worked.

**Table A.13. Number of Full-Time Principals Listed in the Administrative Data and the Number of Schools in Which They Worked, Cohort 1**

| | Treatment | Control |
|---|---|---|
| **Principals Included in the Analyses of Principal Outcomes** | | |
| Year 1 (2011–2012) | | |
| All Principals at the Beginning of the Year | 67 | 70 |
| Full-Time Principals at the Beginning of the Year (*Eligible to be Included in Analysis*) | 65 | 69 |
| Year 2 (2012–2013) | | |
| All Principals at the Beginning of the Year | 69 | 71 |
| Full-Time Principals at the Beginning of the Year (*Eligible to be Included in Analysis*) | 68 | 70 |
| **Schools Included in the Analyses of Principal Outcomes** | | |
| Year 1 (2011–2012) | | |
| All Cohort 1 Schools | 66 | 66 |
| Schools with Principals at the Beginning of the Year | 65 | 66 |
| Schools with Full-Time Principals at the Beginning of the Year | 63 | 65 |
| Year 2 (2012–2013) | | |
| All Cohort 1 Schools | 66 | 66 |
| Schools with Principals at the Beginning of the Year | 66 | 65 |
| Schools with Full-Time Principals at the Beginning of the Year | 65 | 64 |

Source: Educator administrative data.

Note: The number of principals in the analysis might differ from the total number of schools because a few schools did not have a full-time principal or had more than one full-time principal.

We assessed educator effectiveness using several districts' measures used to evaluate and determine TIF performance bonuses, including classroom observation ratings and achievement growth ratings. Table A.14 (teachers) and Table A.15 (principals) describe the sample sizes using different measures of educator effectiveness. In Years 1 and 2, all 132 Cohort 1 schools provided classroom observations ratings for at least some teachers. One district (with 20 schools) did not provide principal observation ratings for Year 1; all 10 Cohort 1 districts provided principal observation ratings for Year 2. Not all schools within a district, however, provided principal observation ratings.

**Table A.14. Teachers Who Had Performance Ratings, Cohort 1 (Percentages)**

| | Treatment | Control | Difference | *p*-value | **Number of Teachers** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Year 1** | | | | | | |
| Had Classroom Observation Rating | 86 | 86 | 0 | 0.754 | **4,346** | **132** |
| Had Classroom Achievement Growth Rating[a] | 38 | 39 | -1 | 0.401 | **2,888** | **73** |
| **Year 2** | | | | | | |
| Had Classroom Observation Rating | 84 | 83 | 1 | 0.583 | **4,466** | **132** |
| Had Classroom Achievement Growth Rating[a] | 44 | 43 | 1 | 0.688 | **2,958** | **73** |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Percentages are based only on teachers in the 6 of 10 districts that evaluated teachers using classroom achievement growth.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.15. Principals Who Had Observation Ratings, Cohort 1 (Percentages)**

| Outcome | Treatment | Control | Difference | *p*-value | **Number of Principals** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Year 1** | | | | | | |
| Had Observation Rating[a] | 100 | 95 | 5 | 0.173 | **108** | **108** |
| **Year 2** | | | | | | |
| Had Observation Rating | 92 | 85 | 7* | 0.043 | **138** | **129** |

Source: Educator administrative data.

Notes: The number of principals exceeds the number of schools in the analysis sample because a few schools had more than one principal. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Percentages are based on 9 of 10 districts that provided data on observation scores for both treatment and control principals in Year 1.

*Difference is statistically significant at the .05 level, two-tailed test.

To help contextualize our findings, in Chapter II, we examined the extent to which educators who received a rating score (and thus were included in the analyses of educator effectiveness) are different from those who did not. We also assessed whether there were differences in the characteristics of treatment and control educators who received ratings. Tables A.16 through A.21 present these findings for the teacher and principal analyses samples. Table A.18 compares characteristics of principals with and without observation ratings in Year 2 only, due to the small number of principals in Year 1 who did not receive an observation rating. Analyses for Tables A.17 and A.20 are based only on teachers in the 6 of 10 districts that evaluated teachers using classroom achievement growth.

**Table A.16. Characteristics of Teachers with and Without Classroom Observation Ratings, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| | Teachers with Observation Ratings | Teachers Without Observation Ratings | Teachers with Observation Ratings | Teachers Without Observation Ratings |
| Demographic Characteristics | | | | |
| Female | 83 | 80 | 85 | 82 |
| Race/ethnicity | | | | |
| White, non-Hispanic | 67 | 66 | 66 | 66 |
| Black, non-Hispanic | 28 | 29 | 29 | 29 |
| Hispanic | 2 | 3 | 2 | 2 |
| Other | 3 | 2 | 3 | 2* |
| Age (average years) | 41 | 41 | 41 | 42 |
| Education | | | | |
| Master's degree or higher | 41 | 43 | 42 | 44 |
| Total experience in K–12 education (average years) | 10 | 10 | 11 | 11 |
| Less than 5 years | 33 | 37 | 32 | 32 |
| 5-15 years | 45 | 38* | 44 | 39* |
| More than 15 years | 22 | 25 | 25 | 29 |
| **Number of Teachers—Range[a]** | **2,586-3,501** | **372-630** | **2,778-3,589** | **373-759** |
| **Number of Schools—Range[a]** | **98-132** | **65-100** | **100-132** | **72-108** |

Source: Educator administrative data.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.17. Characteristics of Teachers with and Without Classroom Achievement Growth Ratings, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| | Teachers with Classroom Achievement Growth Ratings | Teachers Without Classroom Achievement Growth Ratings | Teachers with Classroom Achievement Growth Ratings | Teachers Without Classroom Achievement Growth Ratings |
| Demographic Characteristics | | | | |
| Female | 86 | 83 | 86 | 84 |
| Race/ethnicity | | | | |
| White, non-Hispanic | 63 | 65 | 64 | 65 |
| Black, non-Hispanic | 32 | 28 | 28 | 28 |
| Hispanic | 3 | 4 | 5 | 4 |
| Other | 2 | 3 | 2 | 3 |
| Age (average years) | 39 | 40 | 38 | 41* |
| Education | | | | |
| Master's degree or higher | 36 | 39 | 38 | 40 |
| Total experience in K–12 education (average years) | 9 | 10 | 8 | 10* |
| Less than 5 years | 35 | 35 | 39 | 35 |
| 5-15 years | 46 | 41* | 47 | 43* |
| More than 15 years | 19 | 24 | 14 | 22* |
| **Number of Teachers—Range[a]** | **631-992** | **1,340-1,690** | **937-1,316** | **1,211-1,532** |
| **Number of Schools—Range[a]** | **56-73** | **56-73** | **59-73** | **59-73** |

Source: Educator administrative data.

Note: Analyses are based on 6 of the 10 districts that evaluated teachers using classroom achievement growth.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.18. Characteristics of Principals with and Without Observation Ratings in Year 2, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Principals with Observation Ratings | Principals Without Observation Ratings |
|---|---|---|
| Demographic Characteristics | | |
| Female | 75 | 74 |
| Race/ethnicity | | |
| White, non-Hispanic | 55 | 59 |
| Black, non-Hispanic | 40 | 41 |
| Hispanic | 0 | 0 |
| Other | 5 | 0 |
| Age (average years) | 47 | 53 |
| Education | | |
| Master's degree or higher | 93 | 100 |
| Total experience in K–12 education (average years) | 13 | 14 |
| Less than 5 years | 23 | 35 |
| 5-15 years | 48 | 13* |
| More than 15 years | 30 | 52 |
| **Number of Principals—Range[a]** | **83-111** | **12-19** |
| **Number of Schools—Range[a]** | **82-110** | **12-16** |

Source: Educator administrative data.

Note: The number of principals exceeds the number of schools in the analysis sample because a few schools had more than one principal. Findings for Year 1 are suppressed due to small sample sizes of principals without observation ratings.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.19. Characteristics of Teachers with Classroom Observation Ratings, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 85 | 84 | 1 | 86 | 84 | 2* |
| Race/ethnicity | | | | | | |
| White, non-Hispanic | 74 | 72 | 2 | 73 | 72 | 2 |
| Black, non-Hispanic | 19 | 21 | -2 | 19 | 22 | -3 |
| Hispanic | 2 | 2 | 0 | 3 | 2 | 0 |
| Other | 4 | 4 | 0 | 5 | 4 | 1 |
| Age (average years) | 42 | 41 | 0 | 42 | 41 | 0 |
| Education | | | | | | |
| Master's degree or higher | 51 | 49 | 1 | 49 | 51 | -2 |
| Total experience in K–12 education (average years) | 12 | 11 | 0 | 11 | 11 | 0 |
| Less than 5 years | 23 | 25 | -2 | 27 | 29 | -2 |
| 5-15 years | 47 | 47 | 0 | 46 | 45 | 1 |
| More than 15 years | 30 | 28 | 2 | 27 | 27 | 1 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.285 | | | 0.001 |
| **Number of Teachers—Range[a]** | **1,268-1,763** | **1,318-1,738** | | **1,344-1,781** | **1,434-1,810** | |
| **Number of Schools—Range[a]** | **49-66** | **49-66** | | **50-66** | **50-66** | |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.20. Characteristics of Teachers with Classroom Achievement Growth Ratings, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 87 | 86 | 1 | 88 | 86 | 2 |
| Race/ethnicity | | | | | | |
| White, non-Hispanic | 63 | 63 | 0 | 64 | 62 | 2 |
| Black, non-Hispanic | 31 | 31 | 0 | 29 | 32 | -3 |
| Hispanic | 4 | 3 | 1 | 5 | 4 | 2 |
| Other | 1 | 2 | -1* | 2 | 2 | -1 |
| Age (average years) | 40 | 38 | 1* | 40 | 38 | 1* |
| Education | | | | | | |
| Master's degree or higher | 36 | 38 | -2 | 37 | 37 | 0 |
| Total experience in K–12 | | | | | | |
| education (average years) | 10 | 8 | 2* | 9 | 8 | 2* |
| Less than 5 years | 31 | 36 | -5* | 35 | 41 | -6* |
| 5-15 years | 45 | 48 | -3 | 45 | 48 | -3 |
| More than 15 years | 24 | 16 | 8* | 20 | 11 | 9* |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.000 | | | 0.000 |
| **Number of Teachers—Range[a]** | **299-504** | **332-488** | | **442-644** | **495-672** | |
| **Number of Schools—Range[a]** | **28-37** | **28-36** | | **30-37** | **29-36** | |

Source: Educator administrative data.

Notes: Analyses are based only on teachers in the 6 of 10 districts that evaluated teachers using classroom achievement growth. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

**Table A.21. Characteristics of Principals with Observation Ratings, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 60 | 59 | 1 | 60 | 65 | -5 |
| Race/ethnicity | | | | | | |
| White, non-Hispanic | 64 | 62 | 2 | 59 | 54 | 5 |
| Black, non-Hispanic | 25 | 28 | -3 | 32 | 33 | -1 |
| Hispanic | 4 | 2 | 2 | 4 | 4 | 0 |
| Other | 7 | 8 | -1 | 5 | 9 | -5 |
| Age (average years) | 49 | 48 | 1 | 48 | 48 | 0 |
| Education | | | | | | |
| Master's degree or higher | 93 | 91 | 1 | 96 | 95 | 1 |
| Total experience in K–12 education (average years) | 16 | 15 | 1 | 15 | 14 | 1 |
| Less than 5 years | 15 | 11 | 4 | 18 | 15 | 3 |
| 5-15 years | 38 | 41 | -3 | 37 | 47 | -10 |
| More than 15 years | 47 | 48 | -1 | 45 | 38 | 7 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.000 | | | 0.000 |
| **Number of Principals—Range[a]** | **35-54** | **34-50** | | **44-57** | **39-54** | |
| **Number of Schools—Range[a]** | **35-54** | **34-50** | | **44-57** | **38-53** | |

Source: Educator administrative data.

Notes: The number of principals exceeds the number of schools in the analysis sample because a few schools had more than one principal. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

## Sample Sizes and Analysis of Missing Outcomes in Student Administrative Data

Chapter VI estimates the impact of pay-for-performance on students' math and reading scores on state standardized exams. Table A.22 shows the total number of students with available scores who were the sample for those analyses. Tables A.23 and Table A.24 describe the characteristics of students with and without test scores in math and reading, respectively.

**Table A.22. Students Who Had Test Scores, Cohort 1 (Percentages)**

| | Treatment | Control | Difference | Number of Students | Number of Schools |
|---|---|---|---|---|---|
| **Year 1** | | | | | |
| Math | 93 | 92 | 1 | **44,796** | **132** |
| Reading | 92 | 92 | 0 | **44,796** | **132** |
| **Year 2** | | | | | |
| Math | 92 | 92 | 0 | **44,906** | **132** |
| Reading | 91 | 92 | -1 | **44,906** | **132** |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Table A.23. Characteristics of Students Who Did and Did Not Have Math Test Scores, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| Characteristic | Had Test Scores | Did Not Have Test Scores | Had Test Scores | Did Not Have Test Scores |
| Achievement in Pre-Implementation Year (*z*-score)[a] | | | | |
| Math | -0.44 | -0.79* | -0.44 | -0.80* |
| Reading | -0.39 | -0.71* | -0.39 | -0.69* |
| Race/ethnicity | | | | |
| White, non-Hispanic | 28 | 29 | 28 | 29 |
| African American, non-Hispanic | 42 | 45* | 42 | 45* |
| Hispanic | 24 | 19* | 24 | 20* |
| Other | 6 | 7 | 6 | 7 |
| Other characteristics | | | | |
| Female | 50 | 43* | 49 | 45* |
| Eligible for free/reduced-price lunch | 78 | 79 | 77 | 78 |
| Disabled or has an Individualized Education Program | 12 | 30* | 13 | 34* |
| Overage for grade | 12 | 24* | 12 | 22* |
| Repeating grade | 1 | 2* | 2 | 4* |
| English language learner | 9 | 8 | 7 | 7 |
| Grade Span | | | | |
| Grades 3–5 | 64 | 66 | 64 | 65 |
| Grades 6–8 | 36 | 34 | 36 | 35 |
| **Number of Students—Range[b]** | **23,834-40,886** | **1,506-3,910** | **20,955-40,714** | **1,154-4,192** |
| **Number of Schools—Range[b]** | **84-132** | **80-129** | **84-132** | **72-123** |

Source: Student administrative data.

[a]These averages are only calculated for students who were tested in the pre-implementation year, so they exclude 3rd graders in Year 1 and 3rd and 4th graders in Year 2.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Table A.24. Characteristics of Students Who Did and Did Not Have Reading Test Scores, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | Year 2 | |
|---|---|---|---|---|
| Characteristic | Had Test Scores | Did Not Have Test Scores | Had Test Scores | Did Not Have Test Scores |
| Achievement in Pre-Implementation Year (*z*-score)[a] | | | | |
| Math | -0.43 | -0.83* | -0.44 | -0.81* |
| Reading | -0.38 | -0.72* | -0.39 | -0.73* |
| Race/ethnicity | | | | |
| White, non-Hispanic | 28 | 28 | 28 | 28 |
| African American, non-Hispanic | 43 | 43 | 42 | 44* |
| Hispanic | 23 | 21* | 24 | 22 |
| Other | 6 | 8 | 6 | 7 |
| Other characteristics | | | | |
| Female | 50 | 43* | 50 | 44* |
| Eligible for free/reduced-price lunch | 77 | 79 | 77 | 80 |
| Disabled or has an Individualized Education Program | 12 | 31* | 13 | 33* |
| Overage for grade | 12 | 23* | 12 | 22* |
| Repeating grade | 1 | 2* | 2 | 3* |
| English language learner | 9 | 9 | 7 | 7 |
| Grade Span | | | | |
| Grades 3–5 | 64 | 67 | 64 | 65 |
| Grades 6–8 | 36 | 33 | 36 | 35 |
| **Number of Students—Range[b]** | **23,673-40,592** | **1,554-4,204** | **20,917-40,396** | **1,192-4,510** |
| **Number of Schools—Range[b]** | **84-132** | **81-130** | **84-132** | **80-131** |

Source: Student administrative data.

[a]These averages are only calculated for students who were tested in the pre-implementation year, so they exclude 3rd graders in Year 1 and 3rd and 4th graders in Year 2.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the 0.05 level, two-tailed test.

Our primary analysis in Chapter VI estimates the impact of pay-for-performance on students enrolled in study schools in a given year. As such, our impact estimates measure the impact of pay-for-performance on participating schools, not the impact on individual students. Therefore, this impact can be the result of changes in teacher productivity, changes in teacher composition (due to school mobility), or changes in student composition. Although we cannot disentangle how much of an effect on achievement might result from changes in students or teachers, Tables A.25 and A.26 show that average student characteristics were similar between treatment and control schools across years, suggesting that pay-for-performance did not induce changes in the schools' student composition.

**Table A.25. Characteristics of Students in the Math Analysis Sample, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Characteristic | Treatment | Control | Difference | Treatment | Control | Difference |
| Achievement in the Pre-Implementation Year (*z*-score)[a] | | | | | | |
| Math | -0.46 | -0.41 | -0.05* | -0.46 | -0.43 | -0.03 |
| Reading | -0.40 | -0.38 | -0.02 | -0.40 | -0.40 | 0.00 |
| Race/Ethnicity | | | | | | |
| White, non-Hispanic | 27 | 29 | -2 | 27 | 29 | -2 |
| African American, non-Hispanic | 42 | 42 | 1 | 42 | 41 | 1 |
| Hispanic | 24 | 23 | 2 | 25 | 23 | 2 |
| Other | 6 | 6 | 0 | 6 | 7 | 0 |
| Other Characteristics | | | | | | |
| Female | 49 | 50 | -1 | 49 | 50 | 0 |
| Eligible for free/reduced-price lunch | 77 | 78 | -1 | 77 | 76 | 1 |
| Disabled or has an Individualized Education Program | 12 | 12 | 1 | 13 | 13 | 0 |
| Overage for grade | 12 | 12 | 0 | 12 | 11 | 0 |
| Repeating grade | 1 | 1 | 0 | 2 | 3 | -1* |
| English language learner | 9 | 9 | 0 | 7 | 8 | 0 |
| Grade Span | | | | | | |
| Grades 3–5 | 64 | 64 | 0 | 64 | 64 | -1 |
| Grades 6–8 | 36 | 36 | 0 | 36 | 36 | 1 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.003 | | | 0.054 |
| **Number of Students—Range[b]** | **11,905-20,528** | **11,850-20,324** | | **10,260-20,252** | **10,690-20,457** | |
| **Number of Schools—Range[b]** | **42-66** | **42-66** | | **42-66** | **42-66** | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]These averages are only calculated for students who were tested in the pre-implementation year, so they exclude 3rd graders in Year 1 and 3rd and 4th graders in Year 2.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Table A.26. Characteristics of Students in the Reading Analysis Sample, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Characteristic | Treatment | Control | Difference | Treatment | Control | Difference |
| Achievement in the Pre-Implementation Year (*z*-score)[a] | | | | | | |
| Math | -0.46 | -0.41 | -0.05* | -0.46 | -0.43 | -0.03 |
| Reading | -0.39 | -0.38 | -0.02 | -0.40 | -0.39 | 0.00 |
| Race/Ethnicity | | | | | | |
| White, non-Hispanic | 27 | 30 | -2 | 27 | 29 | -2 |
| African American, non-Hispanic | 43 | 42 | 1 | 42 | 41 | 1 |
| Hispanic | 24 | 23 | 2 | 25 | 23 | 2 |
| Other | 6 | 6 | 0 | 6 | 7 | 0 |
| Other Characteristics | | | | | | |
| Female | 50 | 50 | 0 | 49 | 50 | 0 |
| Eligible for free/reduced-price lunch | 77 | 78 | -1 | 77 | 76 | 1 |
| Disabled or has an Individualized Education Program | 12 | 12 | 0 | 13 | 13 | 0 |
| Overage for grade | 12 | 11 | 0 | 12 | 11 | 0 |
| Repeating grade | 1 | 1 | 0 | 2 | 3 | -1* |
| English language learner | 9 | 9 | 0 | 7 | 8 | 0 |
| Grade Span | | | | | | |
| Grades 3–5 | 64 | 64 | 0 | 64 | 64 | 0 |
| Grades 6–8 | 36 | 36 | 0 | 36 | 36 | 0 |
| Test of whether characteristics jointly predict treatment status: *p*-value | | | 0.009 | | | 0.034 |
| **Number of Students—Range[b]** | **11,804-20,346** | **11,805-20,230** | | **10,220-20,032** | **10,692-20,359** | |
| **Number of Schools—Range[b]** | **42-66** | **42-66** | | **42-66** | **42-66** | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]These averages are only calculated for students who were tested in the pre-implementation year, so they exclude 3rd graders in Year 1 and 3rd and 4th graders in Year 2.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the 0.05 level, two-tailed test.

# APPENDIX B

# SUPPLEMENTAL INFORMATION ON ANALYTIC METHODS FOR CHAPTER II

THIS PAGE IS INTENTIONALLY BLANK

In this appendix, we provide the rationale for and technical details of the methods used in the report. First, we describe how we standardized educator performance ratings and student test scores across districts. Second, we discuss the technical approach for describing the distribution of performance ratings and TIF payouts in evaluation districts. Third, we provide details of the analytic methods used to estimate impacts of pay-for-performance on educator and student outcomes. Fourth, we specify the methods used to impute the maximum pay-for-performance bonus amounts for teachers and principals who reported being eligible for pay-for-performance but who did not answer survey questions about bonus amounts. Fifth, we summarize the level of precision in the study by reporting minimum detectable impacts for key outcomes examined in the impact analyses.

As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts whose schools were randomly assigned in spring and summer 2011 were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. At the time of this report, Cohort 1 had completed two years of implementation, 2011–2012 and 2012–2013, referred to as Years 1 and 2. Cohort 2 districts had completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort.

## Standardizing Outcomes

The two key outcomes discussed in Chapter VI—educator performance ratings and student achievement—were measured using scales or assessments that varied across districts. This section discusses the methods we used to standardize these outcomes for the analysis.

### Educator Performance Ratings

We measured educator effectiveness with several measures that districts used in their TIF programs to evaluate educators and determine performance bonuses. As we noted in Chapter I, districts had to evaluate teachers and principals based on student achievement growth and at least two observations of classroom or school practices. However, districts had flexibility in how they implemented this requirement. For example, districts could choose to evaluate teachers based on the achievement growth of the teachers' own students (classroom achievement growth), all students in the same grade, all students in the school (school achievement growth), or some combination of these measures. Our analysis used four measures: (1) school achievement growth ratings, which were used to evaluate both teachers and principals; (2) teachers' classroom observation ratings; (3) teachers' classroom achievement growth ratings; and (4) principals' observation ratings.

Each of these performance measures either placed educators into three to five performance categories—such as "effective" or "highly effective"—or placed educators onto a numeric scale (typically ranging from 1 to 4 or 1 to 5) in which a one-unit increase was analogous to advancing by a performance level.[74] To express ratings from different districts on a common scale, we transformed the data in two steps. First, if the districts used performance categories but did not already express the

[74] Two districts in Cohort 2 rated educators on a continuous scale for each performance measure and assigned a total score (also on a continuous scale) equal to the sum of the scores from each performance measure. These districts divided the range of the total performance scale (0 to 100) into four intervals, each corresponding to a different performance category. For analysis purposes, we translated educators' scores on each performance measure into the same four categories by dividing the continuous scale of each measure into four intervals, using the same proportional division as the districts used for the total scale.

performance categories as numbers, we ordered the categories and denoted them with consecutive whole numbers, with 1 as the lowest-performing category. This step resulted in all performance ratings being placed on a district-specific numeric scale that had a defined minimum and maximum possible rating. Second, because the range of the scale varied across districts, a one-unit increase would have a different meaning in different districts unless the rating scales were rescaled to have a common range. Therefore, we rescaled all ratings into a common 1-to-4 rating scale with the following formula:

$$(1)\quad \tilde{R}_{jd} = 3 \times \left( \frac{R_{jd} - R_{\min,d}}{R_{\max,d} - R_{\min,d}} \right) + 1$$

where $\tilde{R}_{jd}$ was the rescaled rating of educator *j* in district *d*, $R_{jd}$ was the rating on the district's original numeric scale, and $R_{\min,d}$ $R_{\max,d}$ were the minimum and maximum ratings that educators in district *d* could theoretically receive. Using this formula, an educator who received the lowest rating on the district's scale would receive a rescaled rating of 1, and an educator who received the highest rating on the district's scale would receive a rescaled rating of 4. As another example, an educator who received a 3 on a district scale that ranged from 1 to 5 would have a rescaled rating of 2.5.

At an early stage of the analysis, we explored, but ultimately rejected, an alternative approach to standardizing educator performance ratings across districts. The alternative approach standardized performance ratings into *z*-scores by subtracting district-specific means of the ratings and dividing by district-specific standard deviations of the ratings. We concluded that placing performance ratings on a 1-to-4 scale, as described above, would be preferable to converting the ratings into *z*-scores for several reasons. First, in some districts, estimates of standard deviations would be based on small sample sizes and would therefore not be very reliable. For example, in the smallest evaluation districts that had four to six study schools, only four to six distinct data points would be available for calculating the standard deviation of a school achievement growth rating. Second, some measures produced very little variation in ratings within particular districts, implying that even a small impact (on the original scale) would be misleadingly represented as a huge effect size in *z*-score units. Third, the 1-to-4 rating scale corresponded more closely to the information that educators actually received and to which they would potentially respond.

**Student Achievement**

We measured student achievement with students' scores on state assessments in math and reading. Because student achievement was measured on different scales in different states and grades, we standardized all scores into *z*-scores by subtracting the statewide grade-specific mean and dividing by the statewide grade-specific standard deviation.

We used the following method to eliminate outliers. First, we dropped all scores that were below the minimum or above the maximum values specified by the state assessment's technical manual. Second, we dropped all scores that were more than 5 standard deviations above or below the statewide grade-specific mean. Finally, we recoded scores by giving scores that were between 3.5 and 5 standard deviations above the statewide grade-specific mean the value of 3.5. Similarly, scores that were between -3.5 and -5 standard deviations were given the value of -3.5. Table B.1 shows the percentage of scores that were dropped or recoded, by subject and treatment status. These exclusions and modifications together affected no more than one-half of 1 percent of all scores.

**Table B.1. Test Scores That Were Dropped or Recoded, Cohorts 1 and 2 (Percentages)**

| | Year 1 (Cohorts 1 and 2) | | | Year 2 (Cohort 1) | | |
|---|---|---|---|---|---|---|
| Type of Exclusion or Recoding | Treatment | Control | Difference | Treatment | Control | Difference |
| **Math** | | | | | | |
| Dropped because score was below the minimum score or above the maximum score specified by the technical manual | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | -0.1* |
| Dropped because score was more than 5 standard deviations above or below the statewide mean | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Recoded to 3.5 standard deviations above or below the statewide mean because the score was between 3.5 and 5 standard deviations above or below the statewide mean | 0.2 | 0.2 | 0.1 | 0.2 | 0.2 | 0.0 |
| **Number of Students with Test Scores** | **28,646** | **27,968** | | **20,260** | **20,478** | |
| **Reading** | | | | | | |
| Dropped because score was below the minimum score or above the maximum score specified by the technical manual | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | -0.1* |
| Dropped because score was more than 5 standard deviations above or below the statewide mean | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 |
| Recoded to 3.5 standard deviations above or below the statewide mean because the score was between 3.5 and 5 standard deviations above or below the statewide mean | 0.4 | 0.3 | 0.0 | 0.2 | 0.1 | 0.0 |
| **Number of Students with Test Scores** | **28,350** | **27,781** | | **20,055** | **20,395** | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

*Difference is statistically significant at the 0.05 level, two-tailed test.

## Describing the Average Distribution of Performance Ratings and Payouts

In Chapter IV, we described the distribution—averaged across the 10 Cohort 1 districts—of performance ratings and payouts (including performance bonuses, automatic 1 percent bonuses, and additional pay) that educators received from their TIF programs. We described these distributions with descriptive statistics, including minimum, average, and maximum bonus amounts; percentage of bonus amounts in specific dollar amount ranges; and percentage of performance ratings in specific ranges of the performance scale. Next, we specify how we weighted the data when calculating these descriptive statistics.

We calculated each descriptive statistic in two steps. In the first step, we calculated the descriptive statistic separately within each of the 10 districts. Within each district, we weighted the educator data so that each school contributed equally to the statistic for that district. Specifically, we assigned weights to educators with nonmissing values of the variable so that the sum of their weights was equal across all schools in the district. An educator *j* in school *s* was weighted by weight $W_{js} = 1/N_s$ where $N_s$ was the number of individuals with nonmissing values of the variable in school *s*. In the second step, we took an equal-weighted average of the descriptive statistic across the 10 districts. In supplemental findings (reported in Appendix D), we modified the second step to take a weighted average of the descriptive statistic across the 10 districts, with each district weighted by the number of treatment and control schools in the final analysis sample (see Appendix D, Figures D.3, D.10, D.16, and D.17). Those supplemental findings effectively gave each school the same weight to provide comparable results to the impact analyses, which, as described next, gave equal weight to schools as well.

## Estimating Impacts of Pay-for-Performance on Educator and Student Outcomes

In this section, we describe the estimation model we used to estimate impacts of pay-for-performance on educator and student outcomes, which we presented in Chapters V and VI. We then discuss how we estimated impacts within subgroups defined by educator or student characteristics (presented in Chapters V and VI) or districts' program characteristics (presented in Appendix G) and assessed the differences in impacts between subgroups. For expositional simplicity, we refer primarily to impacts on educator and student outcomes, but we used the same analytic methods to estimates differences between treatment and control schools in educators' understanding and experiences with TIF implementation, which we presented in Chapter IV.

### Main Estimation Model

To estimate the impact of pay-for-performance on educator and student outcomes, we used a regression model that reflected the random assignment design—specifically, the assignment of clusters of educators or students rather than individual educators or students, and the pairing of these clusters before random assignment. We estimated the following model:

$$(2)\quad Y_{js} = \beta T_s + X'_{js}\delta + Z'_s\gamma + W'_s\pi + \varepsilon_{js}$$

where $Y_{js}$ was the outcome for individual (student or educator) *j* in school *s*; $T_s$ was an indicator equal to 1 for treatment schools and zero for control schools; $X_{js}$ was a vector of individual characteristics; $Z_s$ was a vector of school characteristics; $W_s$ was a vector of indicators for the random assignment block (matched pair of schools or matched groups of schools); $\delta$, $\gamma$, and $\pi$ were coefficient vectors

to be estimated; and $\varepsilon_{js}$ was a random error term. The coefficient $\beta$ represented the average impact of pay-for-performance.

We estimated equation (2) using ordinary least squares (OLS) and employed Huber-White sandwich standard errors (Liang and Zeger 1986) to account for the clustering of educator and student outcomes at the level of the random assignment unit (schools or groups of schools). These standard errors were robust to any arbitrary form of correlation among outcomes in the same cluster.[75]

**Covariates**

We controlled for several individual and school covariates in the impact equations to improve precision and adjust for slight preexisting differences between treatment and control schools from the pre-implementation year (2010–2011 for Cohort 1 and 2011–2012 for Cohort 2). For all educator and student outcomes, the school covariates included (1) the school-level averages of math and reading test scores in the pre-implementation year, based on all students in grades 3 to 8 who were tested in the school in the pre-implementation year; and (2) the fractions of the school's enrolled students in grades 3 to 8 who were black, Hispanic, or other race/ethnicity in the pre-implementation year. We chose these covariates because, as shown in Chapter II (Table II.3), there were slight differences between treatment and control schools in average student achievement and racial/ethnic composition in the pre-implementation year.[76]

For some outcomes, we also included individual covariates—those that measured the individual characteristics of educators or students in the analysis samples. These individual covariates allowed for further improvements in precision. The choice of whether to control for individual covariates depended on whether differences in sample composition between treatment and control schools were regarded as random errors (from sampling or random assignment) to be controlled for or whether such differences might actually reflect part of the impact of pay-for-performance. For three categories of outcomes—educators' attitudes, educators' self-reported behaviors, and educator performance ratings—we did not control for individual covariates because pay-for-performance could, in theory, affect those outcomes by way of changing the composition of the educator workforce. For one key outcome, student achievement, and one supplemental outcome, educator retention, we controlled for the characteristics of individuals in the analysis samples, as discussed next.

When estimating impacts on student achievement in Years 1 and 2, we sought to compare students in treatment and control schools who were, on average, equivalent on observed background characteristics. As discussed in Chapter II, we found no evidence that pay-for-performance affected the composition of the student population in the study schools, so we regarded the slight differences in characteristics between students in treatment and control schools as random error to be controlled for. We controlled for students' math and reading test scores from the pre-implementation year; indicators for grade repetition (based on the grade of the assessment), gender, race/ethnicity (indicators for blacks, Hispanics, and students with other race/ethnicity), being old for grade, being an English language learner, having an Individualized Education Program, and receipt of free or reduced-price lunch; interactions between being a grade repeater and baseline test scores; and fixed

[75] As shown in equation (2), we estimated a single average impact from data that were pooled across districts instead of calculating a weighted average of district-specific impacts. This avoided using district-specific estimates whose standard errors could be biased downward due to small numbers of clusters within each district (Donald and Lang 2007).

[76] In the earlier report from this study (Max et al. 2014), we did not control for those covariates because, at the time of writing that report, we had not yet collected districts' administrative data.

effects for combinations of states and assessment grades. Appendix A, Tables A.25 and A.26 show the means of student characteristics (based on nonmissing values) in the math and reading analysis samples, respectively.

In supplemental analyses, we estimated the impact of pay-for-performance on educator retention (Appendix F, Tables F.1 and F.2). Our main measures of educator retention captured whether educators who worked in study schools in Year 1 continued working in the same schools in subsequent years. When estimating impacts on educator retention between Year 1 and subsequent years, we sought to compare treatment and control educators who were, on average, equivalent at the starting point (Year 1) of the analysis period. As Table II.4 shows, treatment and control educators were, indeed, similar in observed characteristics in Year 1, so we regarded any remaining slight differences between the groups as random error to be controlled for. We controlled for dichotomous indicators for gender, race/ethnicity (indicators for whites and blacks), having earned a master's degree or higher, and experience in K–12 education (indicators for 5 to 15 years and more than 15 years), as well as the educator's age in years. Table II.4 shows the means of these variables (based on nonmissing values) in the analysis sample.

**Weights**

We weighted educator and student outcomes so that each school contributed equally to the average impact estimate. Specifically, we assigned weights to individuals with nonmissing outcomes so that the sum of their weights was equal across all schools. An individual *j* in school *s* was weighted by weight $W_{js} = 1/N_s$, where $N_s$ was the number of individuals with nonmissing values for the outcome in school *s*.

**Handling Missing Data**

When estimating impacts on an outcome, our analysis sample included only individuals who had nonmissing values of the outcome variable, and we dropped individuals who had missing values of the outcome variable. Simulations have suggested that, for randomized controlled trials, this approach may have only a small amount of bias (0.05 standard deviations or less) when outcome data are missing at random among individuals with the same covariate values (Puma et al. 2009).

Individuals were not excluded from the analysis samples if they had missing covariate values, as long as they had nonmissing values of the outcome variable. For each covariate, we replaced missing values with a placeholder value (zero). In addition, for each covariate, we constructed an additional binary indicator for whether an individual originally had a missing value for that covariate, and we controlled for this binary indicator in the impact regressions. Simulations by Puma et al. (2009) have shown that this approach to handling missing covariate data is likely to keep estimation bias at less than 0.05 standard deviations.

Tables B.2 through B.5 show the percentages of individuals who were missing covariate values. Although there were some statistically significant differences between treatment and control schools in the percentages of students with missing covariate values, those differences did not exceed 1 percentage point. We found no significant differences in the percentages of teachers or principals in treatment and control schools with missing covariate values, with one exception: treatment principals were more likely than control principals to have missing values for experience in K–12 education.

**Table B.2. Students in the Math Analysis Sample with Missing Covariate Data, Cohort 1 (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Missing Data on: | Treatment | Control | Difference | Treatment | Control | Difference |
| Achievement in the Pre-Implementation Year[a] | | | | | | |
| Math | 34.2 | 33.5 | 0.6 | 57.4 | 56.2 | 1.2 |
| Reading | 34.8 | 34.1 | 0.8 | 57.9 | 56.6 | 1.3 |
| Race/Ethnicity | | | | | | |
| Missing race characteristics | 0.3 | 0.2 | 0.1* | 0.8 | 0.6 | 0.2 |
| Other Characteristics | | | | | | |
| Female | 0.3 | 0.2 | 0.1 | 0.8 | 0.6 | 0.2 |
| Eligible for free/reduced-price lunch | 36.8 | 36.6 | 0.2* | 37.1 | 36.9 | 0.3 |
| Disabled or has an Individualized Education Program | 0.5 | 0.3 | 0.2* | 16.0 | 15.7 | 0.3 |
| Overage for grade | 1.4 | 1.5 | -0.1 | 1.3 | 1.1 | 0.2 |
| Repeating grade | 34.2 | 33.5 | 0.6 | 57.4 | 56.2 | 1.2 |
| English language learner | 0.4 | 0.2 | 0.2* | 16.2 | 15.8 | 0.4 |
| **Number of Students** | **20,528** | **20,324** | | **20,252** | **20,457** | |
| **Number of Schools** | **66** | **66** | | **66** | **66** | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]This characteristic is only defined for students who were tested in the pre-implementation year, so it is missing for 3rd graders in Year 1 and 3rd and 4th graders in Year 2.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Table B.3. Students in the Reading Analysis Sample with Missing Covariate Data, Cohort 1 (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| Missing Data on: | Treatment | Control | Difference | Treatment | Control | Difference |
| Achievement in the Pre-Implementation Year[a] | | | | | | |
| Math | 33.9 | 33.5 | 0.4 | 57.2 | 56.1 | 1.1 |
| Reading | 34.4 | 33.8 | 0.5 | 57.6 | 56.4 | 1.2 |
| Race/Ethnicity | | | | | | |
| Missing race characteristics | 0.3 | 0.2 | 0.1* | 0.8 | 0.6 | 0.2 |
| Other Characteristics | | | | | | |
| Female | 0.3 | 0.2 | 0.1 | 0.7 | 0.6 | 0.1 |
| Eligible for free/reduced-price lunch | 36.8 | 36.6 | 0.2* | 37.1 | 36.9 | 0.2 |
| Disabled or has an Individualized Education Program | 0.5 | 0.3 | 0.2* | 16.0 | 15.8 | 0.2 |
| Overage for grade | 1.4 | 1.5 | -0.1 | 1.3 | 1.1 | 0.1 |
| Repeating grade | 34.3 | 33.7 | 0.5 | 57.6 | 56.3 | 1.3 |
| English language learner | 0.4 | 0.2 | 0.2* | 16.2 | 15.9 | 0.3 |
| **Number of Students** | **20,346** | **20,230** | | **20,032** | **20,359** | |
| **Number of Schools** | **66** | **66** | | **66** | **66** | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]This characteristic is only defined for students who were tested in the pre-implementation year, so it is missing for 3rd graders in Year 1 and 3rd and 4th graders in Year 2.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Table B.4. Teachers in the Educator Retention Analysis Sample with Missing Covariate Data, Cohort 1 (Percentages)**

| Missing Data on: | Treatment | Control | Difference |
|---|---|---|---|
| Sex | 12.3 | 11.7 | 0.6 |
| Race/ethnicity | 4.9 | 4.6 | 0.3 |
| Age | 5.6 | 4.8 | 0.8 |
| Education | 33.0 | 31.9 | 1.1 |
| Experience in K–12 education | 14.5 | 12.5 | 2.0 |
| **Number of Teachers** | **2,189** | **2,157** | |
| **Number of Schools** | **66** | **66** | |

Source: Educator administrative data.

Note: None of the differences between teachers in treatment and control schools were statistically significant at the .05 level. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

**Table B.5. Principals in the Educator Retention Analysis Sample with Missing Covariate Data, Cohort 1 (Percentages)**

| Missing Data on: | Treatment | Control | Difference |
|---|---|---|---|
| Sex | 11.2 | 12.3 | -1.1 |
| Race/ethnicity | 8.3 | 6.2 | 2.1 |
| Age | 9.8 | 8.7 | 1.1 |
| Education | 37.5 | 34.9 | 2.7 |
| Experience in K–12 education | 24.9 | 17.9 | 6.9* |
| **Number of Principals** | **65** | **69** | |
| **Number of Schools** | **63** | **65** | |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

*Difference is statistically significant at the .05 level, two-tailed test.

## Estimation Model for Subgroup Analyses

We estimated the impacts of pay-for-performance within various types of subgroups. In Chapter V, we assessed how the impacts of pay-for-performance on educators' attitudes differed by teachers' teaching assignment and level of experience. In Chapter VI, we examined differences between treatment and control schools in the performance ratings of subgroups of educators defined by their mobility—in particular, educators who stayed in, left, and entered their schools. In Appendix F, we examined the impacts of pay-for-performance on student achievement by grade span. In Appendix G, we assessed how impacts on student achievement differed by districts' program characteristics.

In each type of subgroup analysis, the full sample of students or educators could be partitioned into either two or three mutually exclusive subgroups. For example, suppose that teachers could be partitioned into three subgroups (such as those with low, moderate, and high levels of teaching experience), identified by the binary indicators $Group1_j$, $Group2_j$, and $Group3_j$, respectively. We estimated the following model:

$$(3) \quad \begin{aligned} Y_{js} &= \beta_1 T_s + \gamma_2 Group2_j + \gamma_3 Group3_j + \beta_2 (T_s \times Group2_j) + \beta_3 (T_s \times Group3_j) \\ &+ X_{js}^{'} \delta + Z_s^{'} \gamma + W_s^{'} \pi + \varepsilon_{js}. \end{aligned}$$

In equation (3), the impact of pay-for-performance on teachers in groups 1, 2, and 3 were represented by the parameters $\beta_1$, $(\beta_1 + \beta_2)$, and $(\beta_1 + \beta_3)$. All other variables in equation (3) were the same as those defined in equation (2). We tested the statistical significance of the estimates of $\beta_2$ and $\beta_3$ to determine whether impacts differed across subgroups. For scenarios in which individuals were partitioned into two (rather than three) subgroups, equation (3) was identical except that it did not include indicators and interaction terms involving $Group3_j$.

When examining differences between the performance ratings of treatment and control educators within mobility subgroups, we partitioned the full educator sample in two different ways. First, for analyses of the Year 1 performance ratings of educators who stayed in and left their schools after Year 1, we divided the Year 1 sample of educators into those who stayed in their schools and those who

left their schools after Year 1. Second, for analyses of the Year 2 performance ratings of educators who entered their schools in that year, we divided the Year 2 sample of educators into those who were new to their schools in that year and those who stayed from the previous year.

When examining how impacts varied with a specified program characteristic, our main approach divided districts into two subgroups that differed on that characteristic, allowing us to follow the basic subgroup model shown in equation (3). However, some program characteristics could be expressed as a continuous variable (such as differentiation in teachers' pay-for-performance bonuses and teachers' understanding of their eligibility for pay-for-performance bonuses). For those characteristics, we also estimated a variant of equation (3) that used this continuous measure of the program characteristic. In that model, we did not include indicators, and we replaced the two interaction terms with an interaction between the treatment indicator and the continuous measure of the program characteristic.

**Assessing Variation in Impacts**

To assess the existence of variation in impacts across districts, we estimated a modified version of equation (2) for student achievement outcomes as follows:

$$(4)\quad Y_{js} = \sum_{d=1}^{10} \beta_d (T_s \times I_s^{(d)}) + X_{js}^{'}\delta + Z_s^{'}\gamma + W_s^{'}\pi + \varepsilon_{js}$$

where $I_s^{(d)}$ was an indicator for district *d*, $\beta_d$ represented the impact of pay-for-performance in district *d*, and all other variables were the same as those in equation (2). Equation (4) produced a district-specific impact estimate for each of the 10 Cohort 1 districts. An *F*-test for the joint equality of the 10 impact estimates determined if impacts varied across districts to a statistically significant degree.

**Converting Impacts on School Achievement Growth Ratings into Impacts on Student Test Scores**

In Chapter VI, we reported the impacts of pay-for-performance on districts' measures of school achievement growth. We also reported impacts on student achievement, using administrative data on students' test scores that the study team collected and standardized. Districts used different methods to construct their own measures of school achievement growth from test score data. However, our analysis of impacts on student achievement used the same method for all districts—described earlier in this appendix—to analyze and compare student test scores in treatment and control schools.

To facilitate comparisons between impacts on districts' measures of school achievement growth and impacts on student test scores, we used the following method to convert impacts on school achievement growth ratings (expressed in points on a 1-to-4 rating scale) into implied impacts on student test scores (expressed in student $z$-score units). First, we constructed our own estimate of school achievement growth that was measured directly in student $z$-score units. To construct this measure, we used outcome scores from Year 1 to estimate a modified version of equation (2) that

included baseline individual and school covariates but not a treatment indicator or random assignment block fixed effects:[77]

$$(5)\ Y_{js} = X'_{js}\delta + Z'_{s}\gamma + \varepsilon_{js}.$$

Second, we averaged the predicted residuals $\hat{\varepsilon}_{js}$ from equation (5) to the school level to obtain our estimate of school achievement growth for each subject (math and reading) separately. Third, we calculated the average within-district standard deviation of our school achievement growth measure for each subject, and we took an equal-weighted average of the standard deviations across the two subjects. We found that one standard deviation of our measure of school achievement growth was equal to 0.12 student $z$-score units. Likewise, we calculated the average within-district standard deviation of the districts' school achievement growth ratings, and we found that one standard deviation of the districts' school achievement growth ratings was equal to 0.95 points on the 1-to-4 rating scale. Finally, to convert impacts from one outcome to the other, we assumed that one standard deviation of districts' school achievement growth ratings was equivalent to one standard deviation of our school achievement growth measure.

For example, in Chapter VI, we found that pay-for-performance raised districts' measures of school achievement growth in Year 2 by 0.25 points on the 1-to-4 rating scale (Table VI.2). This impact could be expressed as 0.25 / 0.95 = 0.26 standard deviations of school achievement growth. Assuming that one standard deviation of districts' school achievement growth ratings was equivalent to one standard deviation of our school achievement growth measure, 0.26 standard deviations of school achievement growth could be translated into an implied impact of 0.26*0.12 = 0.03 student $z$-score units.

**Estimating Average Changes in Educator Survey Responses**

We used the following approach to examine whether average educator perceptions of TIF in the study schools changed from Year 1 to Year 2 as bonuses were awarded and educators gained more experience with program components.[78] First, for each school *s* and year *t*, we calculated the average response of educators (indexed by *j*) to the survey item:

$$(6)\ \bar{Y}_{st} = \frac{1}{N_{st}} \sum_{j=1}^{N_{st}} Y_{jst}$$

where $N_{st}$ was the number of educators in school *s* in year *t*. Second, we restricted the sample to schools (indexed from 1 to *N*) that had nonmissing values of $\bar{Y}_{st}$ in both Years 1 and 2. Finally, using both years of data, we estimated the following regression, separately by treatment status:

[77] We did not use Year 2 outcome scores because two years would have elapsed between the baseline assessment (from the pre-implementation year) and the outcome assessment (from Year 2), whereas most school achievement growth measures are based on student growth over the course of one year.

[78] This analysis was not restricted just to teachers who responded to the survey in both years, because such a restriction would not have allowed the analysis to capture changes in average perceptions that resulted from the entry of new teachers in Year 2 who might have had different perceptions than the teachers they replaced.

$$(7)\ \bar{Y}_{st} = \delta Year2_t + \sum_{h=1}^{N} \varphi_h I_s^{(h)} + \omega_{st}$$

where $Year2_t$ was an indicator for Year 2 and $I_s^{(h)}$ was an indicator for school *h*. The coefficient $\delta$ represented the average within-school change in the outcome over time.

## Method for Imputing Missing Values of Educator-Reported Bonus Amounts

For one set of survey items—those that asked educators to report the maximum bonus amounts for which they were eligible—we used a different approach to handling missing data than the approach used for other variables. The reason is that the occurrence of nonresponse in this set of survey items depended upon another variable: whether the educator reported being eligible for the bonus. For simplicity, we refer to a concrete example—teachers' reports of the maximum pay-for-performance bonus amounts for which they were eligible—but the same logic applies to other types of bonuses, as well as to the principal survey. Teachers were asked to report the maximum pay-for-performance bonus amount only if they indicated, in a preceding question, that they were eligible for pay-for-performance. Among teachers who reported being eligible, there was a mix of missing and nonmissing responses to the subsequent question about maximum bonus amounts. On the other hand, among teachers who reported being ineligible, the maximum bonus amount was *always* nonmissing in the analysis because it was defined to be zero.

Consequently, among the full set of teachers who answered the eligibility question, only those who reported being eligible for pay-for-performance could have had a missing report of the maximum bonus amount. This meant that the subset of teachers who had nonmissing values for the maximum bonus amounts was disproportionately made up of teachers who reported being ineligible, and had a maximum bonus amount of zero. Therefore, if only respondents to the bonus amount question were included in the analysis without further corrections for missing data, the average reported maximum bonus amount would have been biased toward zero.

Our solution was to use multiple imputation (MI) to substitute imputed values for missing values of educator-reported bonus amounts among educators who reported being eligible for a specified type of bonus. Because MI accounts for statistical uncertainty in the imputation process, it offers the key analytic advantage of yielding appropriate standard errors for estimates that use the imputed values (Rubin 1987; Schafer and Graham 2002; Puma et al. 2009).

For teachers' reports of maximum bonus amounts, we conducted MI using five steps. First, we pooled data from Years 1 and 2 within Cohort 1 districts and estimated an imputation model in which the reported maximum bonus amount was modeled as a linear function of treatment status, year, treatment status interacted with year, the school covariates listed in the previous section, and random assignment block indicators. We estimated the imputation model using only teachers who reported being eligible for the specified bonus *and* reported a nonmissing bonus amount.[79] Second, we used the estimated coefficients and standard errors from the imputation model to form a posterior distribution

[79] We did not estimate the imputation model separately for the treatment and control groups because this approach would have led to small numbers of teachers per randomization block, resulting in highly imprecise estimates of the coefficients in the imputation model. For imputing a covariate in an analysis model, Puma et al. (2009) advocate estimating imputation models separately by treatment status to avoid artificially creating a correlation between treatment status and the covariate. However, this logic does not apply to imputing a dependent variable of the analysis model, which is the scenario considered here.

for the true coefficients of the imputation model. We made a random draw from this posterior distribution, producing a specific set of coefficients. Third, we used the specific set of coefficients drawn in the previous step to generate predicted values of the perceived bonus amount for all teachers who answered the eligibility question, including respondents and nonrespondents to the question about bonus amounts. Fourth, for each nonrespondent to the bonus amount question, we identified the three respondents who had the closest predicted values to that of the nonrespondent. Fifth, we randomly selected one of these three respondents, and the reported maximum bonus amount of the selected respondent served as the imputed value for the nonrespondent.[80]

We repeated the second through fifth steps 40 times to generate 40 imputed values for each missing value of a teacher-reported bonus amount among teachers who reported being eligible for the specified bonus. We then used these imputed values along with the original, nonmissing values of reported bonus amounts to estimate the analysis model, equation (2), on the full set of teachers who answered the eligibility question. Following standard procedures, we used Rubin's (1987) rules for calculating standard errors of the estimated coefficients in equation (2).

We used the same approach to impute principal-reported maximum bonus amounts. However, unlike for teachers, we did not control for random assignment block indicators in the imputation model due to the small number of principal respondents per block. Instead, we controlled for district indicators.

## Minimum Detectable Impacts

The impact estimation methods described earlier in this appendix were intended, in part, to maximize the precision of the impact estimates. To summarize the level of precision in this study, Table B.6 shows, for each key outcome in this study, the realized value of the minimum detectable impact (MDI) based on the study's actual data, sample definitions, and estimation approach. The MDI was the smallest true impact for which the study had an 80 percent probability of obtaining an estimate that was statistically significant at the 5 percent level. For each outcome, we calculated the MDI as 2.8 multiplied by the standard error of the impact estimate.

[80] Steps 2 through 5 are known as predictive mean matching. In this method, there are no clear rules for choosing the number of respondents with whom a nonrespondent should be matched in step 4. Schenker and Taylor (1996) found that matching each nonrespondent with three respondents performed well in simulations. We followed this approach.

**Table B.6. Realized Values of Minimum Detectable Impacts**

| Outcome | Units | Minimum Detectable Impact |
|---|---|---|
| School Achievement Growth Ratings, Year 1 | Points on 1-to-4 scale | 0.47 |
| School Achievement Growth Ratings, Year 2 | Points on 1-to-4 scale | 0.34 |
| Teachers' Classroom Observation Ratings, Year 1 | Points on 1-to-4 scale | 0.07 |
| Teachers' Classroom Observation Ratings, Year 2 | Points on 1-to-4 scale | 0.08 |
| Teachers' Classroom Achievement Growth Ratings, Year 1 | Points on 1-to-4 scale | 0.22 |
| Teachers' Classroom Achievement Growth Ratings, Year 2 | Points on 1-to-4 scale | 0.15 |
| Observation Ratings for Principals, Year 1 | Points on 1-to-4 scale | 0.22 |
| Observation Ratings for Principals, Year 2 | Points on 1-to-4 scale | 0.27 |
| Student Math Achievement, Year 1 | Student *z*-score units | 0.06 |
| Student Math Achievement, Year 2 | Student *z*-score units | 0.06 |
| Student Reading Achievement, Year 1 | Student *z*-score units | 0.04 |
| Student Reading Achievement, Year 2 | Student *z*-score units | 0.04 |

Source: Educator and student administrative data.

# APPENDIX C

# SUPPLEMENTAL FINDINGS ON PROGRAMS AND EXPERIENCES OF ALL TIF DISTRICTS FOR CHAPTER III

THIS PAGE IS INTENTIONALLY BLANK

This appendix supplements the findings presented in Chapter III and includes additional analyses on TIF districts' programs and challenges implementing TIF. As explained in Chapter II, the final sample for these analyses consisted of 155 TIF districts—13 evaluation and 142 non-evaluation districts—that participated in TIF in 2012–2013 and responded to the 2013 district survey. The 2012–2013 school year, which we refer to as Year 2, was the second year of implementation for nearly all those districts. We refer to the 2011–2012 school year as Year 1.

## TIF Districts and Their Programs

In this section, we provide more details on the measures of educator effectiveness and additional pay opportunities for teachers and principals among all TIF districts. Table C.1 shows additional information on classroom observations for teachers and observations of school practices for principals, as reported by TIF district staff. Table C.2 presents additional pay opportunities for extra work or responsibilities (such as working in a hard-to-staff school) that were not discussed in detail in Chapter III.

**Table C.1. Observations of Classroom or School Practices to Evaluate Teachers and Principals, Year 2 (Percentages Unless Otherwise Noted)**

| | All TIF Districts |
|---|---|
| **Teachers** | |
| Average number of classroom observations per school year | 3.5 |
| Average length of classroom observations (in minutes) | 45 |
| Conducting observations by a trained observer | 99 |
| Classroom observations are conducted by: | |
| Principal or other administrators at the teacher's school | 97 |
| Teacher leaders or peer observers[a] | 57 |
| District administrative staff | 50 |
| Externally hired observers (Non-district employees) | 9 |
| **Number of Districts—Range[b]** | **150-151** |
| **Principals** | |
| Average number of observations per school year | 2.9 |
| Average length of observations (in minutes) | 54 |
| Observations are conducted by: | |
| Superintendent | 50 |
| Other central office administrator from the same district | 58 |
| Administrator from another district | 1 |
| **Number of Districts—Range[b]** | **147-151** |

Source: District survey, 2013.

[a]Department heads, coaches, other senior teachers (at or outside school).

[b]Sample sizes are presented as a range based on the data available for each row in the table.

**Table C.2. Additional Pay Opportunities for Teachers and Principals for Additional Factors, Year 2**

| | Percentage of TIF Districts That Offered Additional Pay | Average Maximum Amount of Additional Pay in Districts Offering it |
|---|---|---|
| **Teachers** | | |
| Additional factors | | |
| Teaching in a hard-to-staff school or high-need subject area | 39 | $3,122 |
| Attending professional development activities or enrolling in graduate-level courses | 32 | $766 |
| **Number of Districts—Range[a]** | **123-154** | **20-137** |
| **Principals** | | |
| Additional factors | | |
| Working in a hard-to-staff school | 14 | $6,279 |
| Attending professional development activities or enrolling in graduate-level courses | 19 | $1,361 |
| **Number of Districts—Range[a]** | **154-155** | **21-58** |

Source: District survey, 2013.

Note: Table reports on activities funded by TIF.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

## Challenges in Implementing TIF

This section provides additional detail on the findings presented in Chapter III on challenges TIF districts faced implementing TIF and revisions they made to their programs. Table C.3 presents results about potential challenges districts faced by whether survey respondents indicated an activity was a "major challenge," "minor challenge," or "not a challenge." Tables C.4 and C.5 show the results for districts' reported revisions to their programs that were discussed in Chapter III.

**Table C.3. Challenges Implementing TIF, Year 2 (Percentages)**

| Activity | Percentage of All TIF Districts Reporting Activity Was: Major Challenge | Minor Challenge | Not a Challenge |
|---|---|---|---|
| Incorporating student achievement growth into teacher evaluations | | | |
| Calculating student achievement growth | 28 | 25 | 47 |
| Attributing student achievement growth to individual teachers | 30 | 29 | 41 |
| Explaining student achievement measures to educators | 28 | 47 | 26 |
| Providing useful and timely feedback on student achievement measures to educators | 33 | 41 | 27 |
| Collecting and storing data linking teachers to student achievement data | 22 | 40 | 38 |
| Teacher classroom observations | | | |
| Choosing a classroom observation tool | 7 | 20 | 73 |
| Finding a tool that is ready for implementation | 9 | 18 | 74 |
| Hiring observers | 2 | 20 | 78 |
| Training observers to use the tool | 10 | 46 | 44 |
| Scheduling and/or conducting observations | 25 | 49 | 26 |
| Providing useful and/or timely feedback from observations | 25 | 46 | 29 |
| Collecting and storing observation data | 13 | 36 | 51 |
| Principal observations | | | |
| Choosing a principal observation tool | 15 | 34 | 51 |
| Finding a tool that is ready for implementation | 17 | 29 | 55 |
| Hiring observers | 1 | 16 | 83 |
| Training observers to use the tool | 7 | 38 | 55 |
| Scheduling and/or conducting observations | 13 | 47 | 40 |
| Providing useful and/or timely feedback from observations | 15 | 40 | 45 |
| Pay-for-performance bonuses | | | |
| Defining the criteria for earning a pay-for-performance bonus or the amount of the bonus | 24 | 41 | 36 |
| Calculating pay-for-performance bonuses | 19 | 31 | 50 |
| Distributing pay-for-performance bonuses | 9 | 35 | 56 |
| Communicating the TIF program to educators or other stakeholders | | | |
| Communicating the TIF program to educators | 15 | 48 | 37 |
| Communicating bonus payouts to educators | 15 | 44 | 41 |
| Communicating with other stakeholders | 13 | 52 | 35 |
| Obtaining or maintaining support for the TIF program | | | |
| Teachers or teachers' union or association | 12 | 31 | 58 |
| Principals or principals' union or association | 1 | 19 | 80 |
| Superintendent | 3 | 14 | 84 |
| School board | 1 | 27 | 72 |
| Parents or broader community | 2 | 25 | 73 |
| Other TIF issues | | | |
| Choosing educators for additional roles and responsibilities | 7 | 47 | 46 |
| Sustainability of the TIF program | 65 | 28 | 7 |
| **Number of Districts—Range[a]** | **147-155** | **147-155** | **147-155** |

Source: District survey, 2013.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

**Table C.4. Revisions to Pay-for-Performance Bonuses After Year 1**

| | All TIF Districts |
|---|---|
| Any revisions to TIF program to change any aspect of pay-for-performance bonuses | 35 |
| Changed pay-for-performance performance assessment or evaluation criteria | 32 |
| Shrank pay-for-performance eligibility | 4 |
| Expanded pay-for-performance eligibility | 13 |
| More teachers were likely to earn a pay-for-performance bonus | 9 |
| Fewer teachers were likely to earn a pay-for-performance bonus | 7 |
| Increased the difference between average and maximum bonus | 6 |
| Decreased the difference between average and maximum bonus | 3 |
| **Number of Districts—Range[a]** | **150-152** |

Source: District survey, 2013.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

**Table C.5. Reasons for Revising the TIF Program After Year 1 to Change Pay-for-Performance Bonuses (Percentages)**

| | All TIF Districts |
|---|---|
| To obtain or maintain support from stakeholders | 14 |
| Principals | 7 |
| Teacher union or association | 6 |
| Teachers | 12 |
| To stay within budget constraints | 11 |
| To simplify the criteria for earning a bonus | 9 |
| To improve perceived fairness | 20 |
| At the request of the U.S. Department of Education | 11 |
| **Number of Districts—Range[a]** | **151-152** |

Source: District survey, 2013.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

# APPENDIX D

# SUPPLEMENTAL FINDINGS ON TIF IMPLEMENTATION IN EVALUATION DISTRICTS FOR CHAPTER IV

THIS PAGE IS INTENTIONALLY BLANK

This appendix supplements the findings presented in Chapter IV on TIF implementation in the evaluation districts. We first provide additional details about the four required components; we then provide additional details on educators' reports about their understanding of the TIF program.

As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts whose schools were randomly assigned in spring and summer 2011 were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 1 districts completed two years of implementation during the period covered by this report, 2011–2012 and 2012–2013, referred to as Years 1 and 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort.

The analyses in Chapter IV were based on Cohort 1 only and, in general, focused on findings for Year 2. This appendix supplements the findings in Chapter IV in several ways: (1) we present findings for Cohort 1 that were noted but not included in the chapter, such as findings for Year 1; (2) we provide findings for Year 1 based on Cohorts 1 and 2; (3) we show findings when we weight data on pay-for-performance bonuses by the number of schools in a district, rather than giving each district equal weight; and (4) we present findings from subgroup analyses to examine factors that might explain differences in teachers' understanding of their bonus eligibility.

## Implementation of the Required Components of TIF

In this section, we show results presented in Chapter IV about the components of TIF programs that the evaluation districts designed and implemented, focusing on the four required components under the TIF grant: (1) measures of educator effectiveness, (2) pay-for-performance bonuses, (3) additional pay opportunities, and (4) professional development.

### Requirement 1: Measures of Educator Effectiveness

TIF grantees were required to measure educator effectiveness based on student achievement growth and multiple observations by trained observers. Chapter IV focused on Cohort 1 districts' implementation of this requirement in Year 2. Table D.1 shows additional details on teacher classroom observations, as reported by the Cohort 1 districts. Figure IV.1 in Chapter IV illustrated the distribution of teacher performance ratings for Cohort 1 in Year 2, based on administrative data. Figure D.1 shows the distribution of teacher performance ratings for Cohort 1 in Year 1. Similar to the Year 2 findings, most teachers in Year 1 (68 percent) were rated in the top two quarters of the rating scale on classroom observations, but many fewer (24 percent) were rated in the top two quarters of the rating scale for school-level student achievement growth. Figure D.2 shows the distribution of teacher performance ratings in Year 1 for Cohorts 1 and 2. Similar to the findings for Cohort 1 only, most teachers (69 percent) in Cohorts 1 and 2 were rated in the top two quarters of the rating scale on classroom observations, and fewer than 25 percent were rated in the top two quarters for school-level student achievement growth. Table D.2 shows the distribution of principal performance ratings for Cohort 1 in Years 1 and 2. Most principals (62 percent in Year 1 and 72 percent in Year 2) were rated in the top two quarters of the rating scale for observations, and fewer than one-fourth were rated in the top two quarters for school achievement growth.

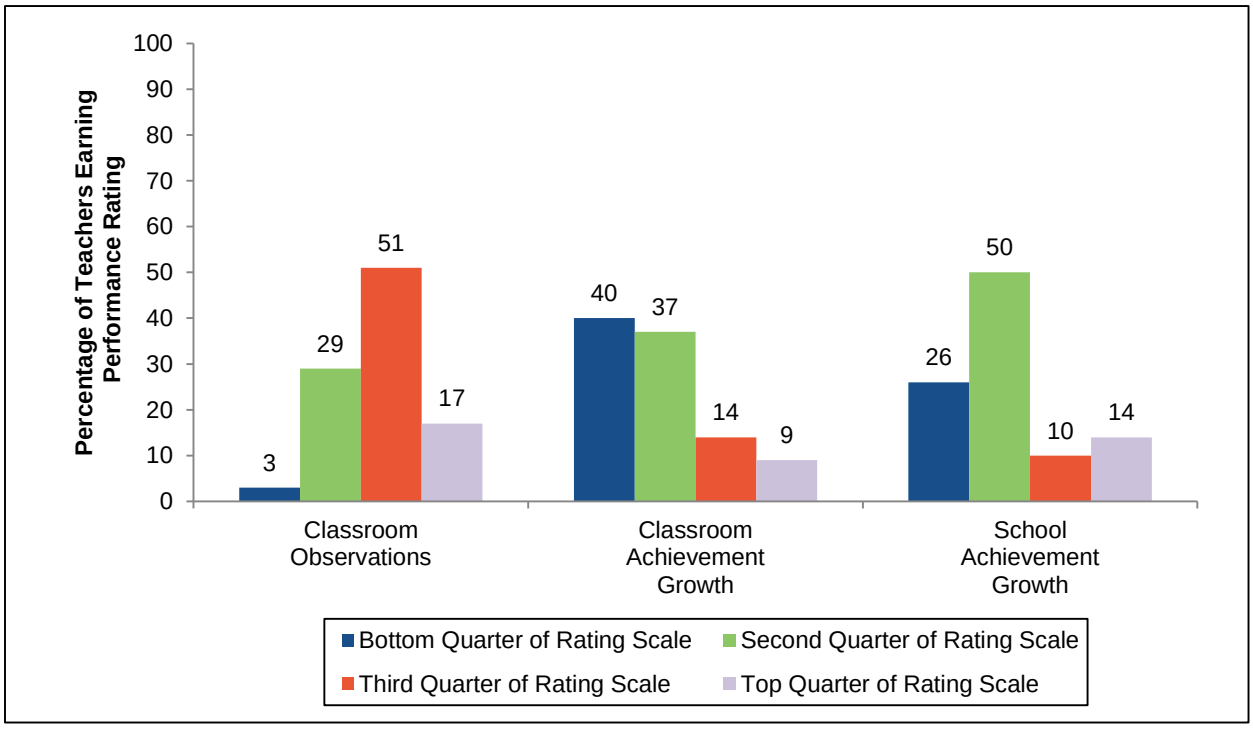
**Table D.1. Classroom Observations to Evaluate Teachers in Year 2, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Evaluation Districts |
|---|---|
| Average number of observations per school year | 3.5 |
| Average length of observations (in minutes) | 47 |
| Conducting observations by a trained observer | 100 |
| Observations are conducted by: | |
| Principal or other administrators at the teacher's school | 90 |
| Teacher leaders or peer observers[a] | 50 |
| District administrative staff | 40 |
| Externally hired observers (nondistrict employees) | 10 |
| **Number of Districts** | **10** |

Source: District survey, 2013.

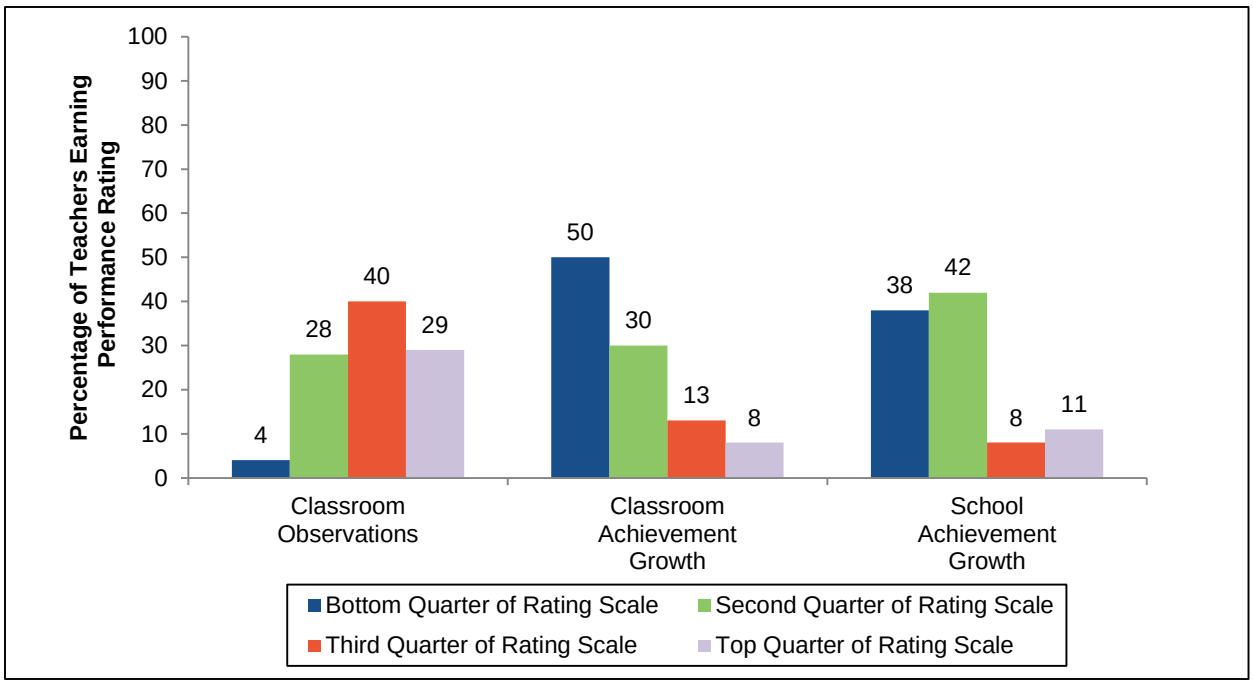
[a]Department heads, coaches, other senior teachers (at or outside school).

**Figure D.1. Distribution of Teachers' Performance Ratings in Year 1, Cohort 1**




Source: Educator administrative data (N = 3,625 teachers for the classroom observation score rating, N = 1,093 teachers for the classroom student achievement growth rating, and N = 4,186 teachers for the school student achievement growth rating).

**Figure D.2. Distribution of Teachers' Performance Ratings in Year 1, Cohorts 1 and 2**




Source: Educator administrative data (N = 5,219 teachers for the classroom observation score rating, N = 2,439 teachers for the classroom student achievement growth rating, and N = 5,231 teachers for the school student achievement growth rating).

**Table D.2. Distribution of Principal Performance Ratings, Cohort 1**

| Year and Performance Measure | Percentage of Principals Earning Performance Ratings in Specified Portion of Rating Scale | | | | Number of Principals | Number of Schools |
|---|---|---|---|---|---|---|
| | Bottom Quarter of Scale | Second Quarter of Scale | Third Quarter of Scale | Top Quarter of Scale | | |
| **Year 1** | | | | | | |
| Observation Score | 8 | 20 | 45 | 27 | 111 | 111 |
| School Achievement Growth | 26 | 50 | 10 | 14 | 127 | 121 |
| **Year 2** | | | | | | |
| Observation Score | 1 | 17 | 55 | 27 | 118 | 117 |
| School Achievement Growth | 25 | 52 | 9 | 13 | 137 | 128 |

Source: Educator administrative data.

In Chapter IV, Figure IV.2 illustrated that school achievement growth and classroom observations sometimes identified the same teachers as high-performing in Year 2, but many teachers had higher ratings from observations than from school achievement growth. Table D.3 shows the degree of consistency between school achievement growth and classroom observations in Year 1 for Cohort 1. In Year 1, as in Year 2, most teachers (70 percent) who scored low (in the bottom quarter of the rating scale) on school achievement growth scored at least moderately high (in the top half of the rating scale) on classroom observations. However, whereas in Year 2 teachers who scored high (in the top quarter) on school achievement growth were more likely to score high on observations than teachers who scored low on school achievement growth, in Year 1 about similar percentages (17 or 18 percent) of teachers who scored low and high on school achievement growth had high observation ratings.

Tables D.4 and D.5 show the degree of consistency between classroom achievement growth and classroom observations for teachers in Cohort 1 within Years 1 and 2, respectively. No more than 10 percent of teachers with low classroom achievement growth ratings had high observation ratings. However, many teachers (53 percent in Year 1 and 47 percent in Year 2) who scored low on classroom achievement growth had at least moderately high observation ratings.

Tables D.6 and D.7 show the degree of consistency between school achievement growth and observations for principals in Cohort 1 within Years 1 and 2, respectively. Many principals (47 percent in Year 1 and 72 percent in Year 2) with low school achievement growth ratings had at least moderately high observation ratings.

**Table D.3. Degree of Consistency Between School Achievement Growth and Classroom Observations for Teachers in Year 1, Cohort 1**

| | Percentage of Teachers Earning Classroom Observation Ratings in Specified Portion of Scale | | | | | |
|---|---|---|---|---|---|---|
| Portion of Scale in which Teacher Earned School Achievement Growth Rating | Bottom Quarter of Observation Scale | Second Quarter of Observation Scale | Third Quarter of Observation Scale | Top Quarter of Observation Scale | Total | Number of Teachers |
| Bottom Quarter of Growth Scale | 2 | 28 | 53 | 17 | 100 | 936 |
| Second Quarter of Growth Scale | 5 | 34 | 45 | 16 | 100 | 1,500 |
| Third Quarter of Growth Scale | 2 | 42 | 48 | 8 | 100 | 403 |
| Fourth Quarter of Growth Scale | 1 | 13 | 68 | 18 | 100 | 649 |

Source: Educator administrative data.

**Table D.4. Degree of Consistency Between Classroom Achievement Growth and Classroom Observations for Teachers in Year 1, Cohort 1**

| Portion of Scale in which Teacher Earned Classroom Achievement Growth Rating | Percentage of Teachers Earning Classroom Observation Ratings in Specified Portion of Scale | | | | | |
|---|---|---|---|---|---|---|
| | Bottom Quarter of Observation Scale | Second Quarter of Observation Scale | Third Quarter of Observation Scale | Top Quarter of Observation Scale | Total | Number of Teachers |
| Bottom Quarter of Growth Scale | 5 | 42 | 43 | 10 | 100 | 486 |
| Second Quarter of Growth Scale | 5 | 49 | 42 | 3 | 100 | 320 |
| Third Quarter of Growth Scale | 0 | 31 | 62 | 7 | 100 | 128 |
| Fourth Quarter of Growth Scale | 0 | 17 | 70 | 12 | 100 | 106 |

Source: Educator administrative data.

**Table D.5. Degree of Consistency Between Classroom Achievement Growth and Classroom Observations for Teachers in Year 2, Cohort 1**

| Portion of Scale in which Teacher Earned Classroom Achievement Growth Rating | Percentage of Teachers Earning Classroom Observation Ratings in Specified Portion of Scale | | | | | |
|---|---|---|---|---|---|---|
| | Bottom Quarter of Observation Scale | Second Quarter of Observation Scale | Third Quarter of Observation Scale | Top Quarter of Observation Scale | Total | Number of Teachers |
| Bottom Quarter of Growth Scale | 4 | 49 | 43 | 4 | 100 | 565 |
| Second Quarter of Growth Scale | 2 | 44 | 47 | 7 | 100 | 345 |
| Third Quarter of Growth Scale | 1 | 19 | 56 | 24 | 100 | 179 |
| Fourth Quarter of Growth Scale | 0 | 16 | 50 | 34 | 100 | 225 |

Source: Educator administrative data.

**Table D.6. Degree of Consistency Between School Achievement Growth and Observations for Principals in Year 1, Cohort 1**

| Portion of Scale in which Principal Earned School Achievement Growth Rating | Percentage of Principals Earning Observation Ratings in Specified Portion of Scale | | | | | |
|---|---|---|---|---|---|---|
| | Bottom Quarter of Observation Scale | Second Quarter of Observation Scale | Third Quarter of Observation Scale | Top Quarter of Observation Scale | Total | Number of Principals |
| Bottom Quarter of Growth Scale | 17 | 18 | 28 | 37 | 100 | 30 |
| Second Quarter of Growth Scale | 5 | 26 | 42 | 27 | 100 | 48 |
| Third Quarter of Growth Scale | 0 | 27 | 66 | 7 | 100 | 7 |
| Fourth Quarter of Growth Scale | 13 | 10 | 54 | 23 | 100 | 19 |

Source: Educator administrative data.

**Table D.7. Degree of Consistency Between School Achievement Growth and Observations for Principals in Year 2, Cohort 1**

| Portion of Scale in which Principal Earned School Achievement Growth Rating | Percentage of Principals Earning Observation Ratings in Specified Portion of Scale | | | | | |
|---|---|---|---|---|---|---|
| | Bottom Quarter of Observation Scale | Second Quarter of Observation Scale | Third Quarter of Observation Scale | Top Quarter of Observation Scale | Total | Number of Principals |
| Bottom Quarter of Growth Scale | 2 | 20 | 56 | 22 | 100 | 37 |
| Second Quarter of Growth Scale | 0 | 16 | 57 | 27 | 100 | 55 |
| Third Quarter of Growth Scale | 0 | 21 | 41 | 38 | 100 | 9 |
| Fourth Quarter of Growth Scale | 0 | 12 | 56 | 33 | 100 | 16 |

Source: Educator administrative data.

## Requirement 2: Pay-for-Performance Bonuses

This section presents additional information on districts' pay-for-performance programs and analyses on pay-for-performance bonuses. The additional analyses examine whether the findings change if we (1) base findings for Year 1 on Cohorts 1 and 2 (rather than just Cohort 1), or (2) weight districts by the number of schools (rather than weight each district equally). We also provide information that supports statements in Chapter IV (such as the distribution of bonuses by district) and provide findings for Cohort 1 in Year 1 (or Year 2) when the findings for that year were not provided in Chapter IV. We provide additional information for teachers first, then for principals.

Tables D.8 and D.9 provide additional information on Cohorts 1 and 2 pay-for-performance programs. Table D.8 provides summary information on key features of districts' programs, while Table D.9 provides more detailed information on their programs. To ensure districts' confidentiality, the numbering of the districts in these tables does not mirror the lettering of districts in other parts of the report.

**Table D.8. Key Features of Evaluation Districts' Teacher Pay-for-Performance Bonus Programs in Year 2, Cohorts 1 and 2**

| | Cohort 1 Districts | | | | | | | | | | Cohort 2 Districts | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Key Program Feature | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Teachers could receive a bonus for multiple performance measures | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Teachers could receive a bonus for a single overall performance rating | | | | | | | | | | | | ✓ | ✓ |
| Teachers could receive a bonus for school achievement growth | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Teachers in tested grades and subjects could receive a bonus for their students' achievement growth | | | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | |
| Teachers could receive a bonus for the achievement growth of a student subgroup | | | | | ✓ | ✓ | | | ✓ | | | | |
| Student achievement growth was measured by a value-added model | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teachers could receive a bonus for observations | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | ✓ | ✓ | | |
| A maximum bonus was specified for each performance measure or for overall rating | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| Maximum bonus possible depended on the number of bonus recipients | ✓ | | ✓ | ✓ | | | | | | ✓ | ✓ | | |
| Bonus amount for a performance measure could be affected by a factor besides the teacher's rating on the measure | | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | |
| District changed some aspect of its program between Year 1 and Year 2 | ✓ | ✓ | | | | | ✓ | | ✓ | | NA | NA | NA |

Source: District interviews from 2012 and 2013, grantees' Annual Performance Report (APR) documents, and technical assistance documents.

Note: Grantees submit an APR to the U.S. Department of Education that describes how educators are evaluated.

NA = not applicable.

**Table D.9. Detailed Information on Measures and Criteria Used for Evaluation Districts' Teacher Pay-for-Performance Bonus Programs in Year 2, Cohorts 1 and 2**

**Cohort 1**

**District 1**

**Key program features**

- Teachers could receive bonuses for school achievement growth and classroom observations
- Maximum bonus possible for classroom observations depended on number of bonus recipients; maximum bonus possible for other measures was fixed
- Revised its program between Year 1 and Year 2

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - Based on school value-added score
   - School's 2012–2013 value-added ranking was compared to school's 2011–2012 value-added ranking
   - Maximum bonus received if school met Target 1, defined as the value-added score the school was estimated to have 25 percent probability of achieving based on 2011–2012 performance
   - Smaller bonus received if school met Target 2, defined as the value-added score the school was estimated to have 50 percent probability of achieving based on 2011–2012 performance

2. Bonuses based on classroom observations
   - Teachers were observed 6 times during the year
   - Pool of money set aside for observation bonuses
   - Could receive up to 4 points for each standard on the rubric
   - Awards were based on the total number of points a teacher received
   - The total possible point count was partitioned into 4 tiers
   - Tiers were determined at the end of the school year
   - Teachers received a bonus if their total score fell within the top 3 tiers, and received the maximum bonus if their total score fell in the top tier

**District 2**

**Key program features**

- Teachers could receive bonuses for school achievement growth in math, school achievement growth in ELA, and classroom observations
- Set an absolute maximum bonus possible for each criterion
- Revised its program between Year 1 and Year 2

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth in math
   - Based on school math value-added score
   - School achievement growth was partitioned into 4 tiers: (a) Tier 1: 90-100th percentile, (b) Tier 2: 80-89th percentile, (c) Tier 3: 65-79th percentile; (d) Tier 4: below the 65th percentile
   - Teachers in Tier 4 schools did not receive a bonus
   - The maximum bonus went to teachers in the Tier 1 schools

2. Bonuses based on school achievement growth in ELA
   - Based on school ELA value-added score
   - School achievement growth in ELA was partitioned into 4 tiers: (a) Tier 1: 90-100th percentile, (b) Tier 2: 80-89th percentile, (c) Tier 3: 65-79th percentile; (d) Tier 4: below the 65th percentile
   - Teachers in Tier 4 schools did not received a bonus
   - The maximum bonus went to teachers in the Tier 1 schools

3. Bonuses based on classroom observations
   - Teachers were observed 6 times during the year
   - Scores ranged from 1 to 4
   - Teachers received the maximum bonus if their average score was 3.7 or above and they earned at least a 3 on each evaluation
   - Teachers received the second highest bonus if their average score was between 3.4 and 3.69 and they earned at least a 2 on each evaluation
   - Teachers received the smallest bonus if their average score was between 3.0 and 3.39

**Cohort 1**

**Districts 3 and 4**

**Key program features**

- Teachers could receive bonuses for school achievement growth, classroom achievement growth (if teaching tested grades and subjects), and classroom observations
- For each performance measure, teachers' ratings were translated into "shares" that determined their bonus amounts
- Maximum bonus possible for each measure depended on the number of bonus recipients
- Bonus based on observations depended on a factor besides the observation score

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - For teachers in tested grades and subjects, 20 percent of their potential bonus was based on school achievement growth
   - For teachers in nontested grades and subjects, 50 percent of their potential bonus was based on school achievement growth
   - Based on school value-added score placed onto a 1-to-5 scale
   - Teachers in schools rated 1 or 2 earned 0 shares; teachers in schools rated 3 earned 50 shares; teachers in schools rated 4 earned 75 shares; teachers in schools rated 5 earned 100 shares.

2. Bonuses based on classroom achievement growth
   - For teachers in tested grades and subjects, 30 percent of their potential bonus was based on classroom achievement growth
   - Based on classroom value-added score placed onto a 1-to-5 scale
   - Teachers rated 1 or 2 earned 0 shares; teachers rated 3 earned 1 share; teachers rated 4 earned 6 shares, teachers rated 5 earned 10 shares

3. Bonuses based on classroom observations
   - For all teachers, 50 percent of their potential bonus was based on classroom observations
   - Teachers were observed 4 times during the year
   - Teachers were classified into 1 of 4 possible categories: (1) career teacher, (2) teacher in a hard-to-fill position, (3) mentor teacher, or (4) master teacher
   - The number of shares earned depended on the teacher's observation rating and position
   - Teachers earned more shares the higher their observation score, but had to be rated above a minimum score to receive any shares
   - The minimum observation score required to receive shares varied depending on their position
   - For a given observation rating, career teachers and teachers in a hard-to-fill position earned more shares than mentor or master teachers

**District 5**

**Key program features**

- Teachers could receive bonuses for school achievement growth, grade-level achievement growth, and classroom achievement growth (if teaching tested grades and subjects)
- Set an absolute maximum bonus possible for each criterion
- Teachers could not receive a bonus for classroom observations; however, a teacher's total bonus (based on other measures) was reduced by 25 percent if the teacher's observation score did not meet a minimum threshold
- Bonus based on grade-level achievement growth depended on a factor besides the student subgroups' score

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - Based on school value-added score
   - Bonuses were awarded to teachers in schools whose school value-added score was at least 1 standard error (SE) above the state average

2. Bonuses based on grade-level achievement growth
   - Based on grade-level value-added score
   - All teachers joined a grade-level team

**Cohort 1**

*District 5 (continued)*

- Bonus were awarded to teachers in grades whose grade-level value-added score was at least 1 SE above the state average
- Bonus depended on the percentage of time teacher spent working with that grade

3. Bonuses based on classroom achievement growth
   - Based on classroom value-added score
   - Awards of increasing value were given to teachers whose value added score was at least (1) 0.5 SE above the state average, (2) 1.0 SE above the state average, (3) 1.5 SE above the state average, and (4) 2.0 SE above the state average

**District 6**

**Key program features**

- Teachers could receive bonuses for school achievement growth, achievement growth of student subgroups, and classroom observations
- Set an absolute maximum bonus possible for each criterion
- Bonus based on classroom observations depended on factors besides the observation score

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - Based on Colorado Growth Model
   - Each school set a goal for its Colorado Growth Model score
   - Bonuses were awarded if the school met its goal

2. Bonuses based on achievement growth of student subgroups
   - All teachers were assigned to a team
   - Teams of teachers set goals for the achievement growth of their students
   - Bonuses were awarded if the team met its goal

3. Bonuses based on classroom observations
   - Teachers were observed an average of 3 times per year
   - The size of the bonus depended on the teacher's years of education, highest degree earned, and score on the rubric

**District 7**

**Key program features**

- Teachers could receive bonuses for school achievement growth, classroom achievement growth (if teaching tested grades and subjects), classroom observations, and school achievement levels
- Set an absolute maximum bonus possible for each criterion
- Revised its program between Year 1 and Year 2

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - Fall-to-spring growth targets were set for each student based on the student's fall achievement
   - Schools were rated on a 1 to 4 scale based on how their students' growth compared with the targets
   - Teachers in schools rated 4 earned a bonus worth 2 percent of average teacher salary; teachers in schools rated 3 earned a bonus worth 1.5 percent of average teacher salary; teachers in schools rated 2 earned a bonus worth 1 percent of average teacher salary; teachers in schools rated 1 did not earn a bonus for this measure

2. Bonuses based on classroom achievement growth
   - Bonus for the measure was available for math, science, and ELA teachers only
   - Fall-to-spring growth targets were set for each student based on the student's fall achievement
   - Teachers were rated on a 1 to 4 scale based on how their students' growth compared with the targets
   - Teachers rated 4 earned a bonus worth 5 percent of average teacher salary; teachers rated 3 earned a bonus worth 3.5 percent of average teacher salary; teachers rated 2 earned a bonus worth 1 percent of average teacher salary; teachers rated 1 did not earn a bonus for this measure

3. Bonuses based on classroom observations
   - Bonus awarded for score on third party rating of video of a classroom lesson
   - Teachers were rated on a 1 to 4 scale

**Cohort 1**

*District 7 (continued)*

- For math, science, and ELA teachers, those rated 4 earned a bonus worth 4 percent of average teacher salary; those rated 3 earned a bonus worth 3 percent of average teacher salary; those rated 2 earned a bonus worth 1 percent of average teacher salary; those rated 1 did not earn a bonus for this measure
- For other teachers, those rated 4 earned a bonus worth 6 percent of average teacher salary; those rated 3 earned a bonus worth 4 percent of average teacher salary; those rated 2 earned a bonus worth 1 percent of average teacher salary; those rated 1 did not earn a bonus for this measure

4. Bonuses based on school's achievement level

- Bonus awarded for school's performance score on the state test
- Ratings were put on a 1 to 4 scale
- Teachers in schools rated 4 earned bonus worth 2 percent of average teacher salary; teachers in schools rated 3 earned bonus worth 1.5 percent of average teacher salary; teachers in schools rated 2 earned bonus worth 1 percent of average teacher salary; teachers in schools rated 1 did not earn a bonus for this measure

**District 8**

**Key program features**

- All teachers could receive a bonus for school achievement growth; teachers in tested grades and subjects could also receive a bonus for classroom achievement growth
- Set an absolute maximum bonus possible for each criterion
- Teachers could not receive a bonus for classroom observations; however, a teacher had to be rated at least proficient on the summative observation score to earn a bonus for school or classroom achievement growth

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - Based on school value-added score
   - School must receive a rating of "exceeds expected growth" to receive bonus
   - Schools were rated as "exceeds expected growth" if their value-added score was at least 1 standard deviation above the state mean

2. Bonuses based on classroom achievement growth
   - Bonus available to teachers in tested grades and subjects
   - Based on classroom value-added score
   - Teachers with scores between 1 and 1.9 standard deviations above the mean received a rating of 4; teachers with scores at least 2 standard deviations above the mean received a rating of 5
   - Bonuses awarded to teachers with ratings of 4 or 5
   - Math teachers received larger bonuses than non-math teachers

**District 9**

**Key program features**

- Teachers could receive bonuses for school achievement growth, achievement growth attributable to teacher teams, school achievement levels, and achievement levels attributable to teacher teams
- Set an absolute maximum bonus possible for each criterion
- Teachers could not receive a bonus for classroom observations; however, a teacher had to be rated 3 or above on the summative observation measure to receive bonuses based on other measures
- Teachers had their bonuses prorated if they were in attendance for less than 95 percent of the school year, and could not receive any bonus if they were in attendance for less than 80 percent of the school year
- Revised its program between Year 1 and Year 2

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - Based on school value-added score
   - Teachers in schools whose value-added score was rated above expected growth earned a bonus

2. Bonuses based on achievement growth attributable to teacher teams
   - All teachers joined one of four subject-matter teams: math, ELA, science, or social studies
   - Teachers in a subject-matter team received a bonus if their school's value-added score for the specified subject was rated above expected growth

**Cohort 1**

*District 9 (continued)*

3. Bonuses based on school achievement levels
   - Teachers in schools whose performance index increased by a minimum required amount earned a bonus
   - The minimum required gain in the performance index depended on the school's performance index in the prior year

4. Bonuses based on achievement levels attributable to teacher teams
   - All teachers joined one of four subject-matter teams: math, ELA, science, or social studies
   - Teams set goals for student achievement in their subject
   - Teachers in teams that met their goals received a bonus

**District 10**

**Key program features**

- Teachers could receive bonuses for school achievement growth, classroom achievement growth (if teaching tested grades and subjects), and classroom observations
- Maximum bonus possible for each measure depended on the number of bonus recipients
- Bonus based on classroom observations depended on a factor besides the observation score

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - For teachers in tested grades and subjects, 20 percent of their potential bonus was based on school achievement growth
   - For teachers in nontested grades and subjects, 50 percent of their potential bonus was based on school achievement growth
   - Based on school value-added scores placed on a 1 to 5 scale
   - Teachers in schools rated 3 or higher earned a bonus, with larger bonuses to teachers in schools with higher ratings

2. Bonuses based on classroom achievement growth
   - For teachers in tested grades and subjects, 30 percent of their potential bonus was based on classroom achievement growth
   - Based on classroom value-added scores placed on a 1 to 5 scale
   - Teachers rated 3 or higher earned a bonus, with larger bonuses to teachers with higher ratings

3. Bonuses based on classroom observations
   - For all teachers, 50 percent of their potential bonus was based on classroom observations
   - Teachers were observed 4 times during the year
   - Teachers were classified into 1 of 4 possible positions: (1) career teacher, (2) teacher in a hard-to-fill position, (3) mentor teacher, or (4) master teacher
   - Observation scores were put on a 1 to 5 scale
   - The size of the bonus earned depended on the teacher's observation rating and position
   - Teachers earned larger bonuses the higher their observation rating, but had to be rated at or above a minimum rating to receive a bonus, which depended on their position

**Cohort 2**

**District 11**

**Key program features**

- Teachers could receive bonuses for school achievement growth, classroom achievement growth (if teaching tested grades and subjects), and classroom observations
- For each performance measure, teachers' ratings were translated into "shares" that determined their bonus amounts
- Maximum bonus possible for each measure depended on the number of bonus recipients
- Bonus for classroom observations depended on a factor besides the observation score

**Specific information on performance measures and bonus criteria**

1. Bonuses based on school achievement growth
   - For teachers in tested grades and subjects, 20 percent of their potential bonus was based on school achievement growth
   - For teachers in nontested grades and subjects, 50 percent of their potential bonus was based on school achievement growth
   - Based on school value-added score placed on a 1 to 5 scale

**Cohort 2**

*District 11 (continued)*

- Teachers in schools rated 1 or 2 earned 0 shares; teachers in schools rated 3 earned 50 shares; teachers in schools rated 4 earned 75 shares; teachers in schools rated 5 earned 100 shares

2. Bonuses based on classroom achievement growth
   - For teachers in tested grades and subjects, 30 percent of their potential bonus was based on classroom achievement growth
   - Based on classroom value-added score placed on a 1 to 5 scale
   - Teachers rated 1 or 2 earned 0 shares; teachers rated 3 earned 1 share; teachers rated 4 earned 6 shares, teachers rated 5 earned 10 shares

3. Bonuses based on classroom observations
   - For all teachers, 50 percent of their potential bonus was based on classroom observations
   - Teachers were observed 4 times during the year
   - Teachers were classified into 1 of 4 possible categories: (1) career teacher, (2) teacher in a hard-to-fill position, (3) mentor teacher, or (4) master teacher
   - The number of shares earned depended on the teacher's observation rating and position
   - Teachers earned more shares the higher their observation score, but had to be rated above a minimum score to receive any shares
   - The minimum observation score required to receive shares varied depending on their position
   - For a given observation rating, career teachers and teachers in a hard-to-fill position earned more shares than mentor or master teachers

**District 12**

**Key program features**

- Teachers could receive a bonus for 1 overall performance measure that combined ratings based on classroom achievement growth, classroom observations, and classroom achievement levels
- Set an absolute maximum bonus
- Teachers receiving a score of 4 on a 1 to 4 scale received a bonus
- Only teachers in tested grades and subjects were eligible for bonuses

**Specific information on performance measures**

1. Rating based on classroom achievement growth
   - Based on value-added score on state assessment
   - 20 percent of overall evaluation score based on classroom achievement growth

2. Rating based on classroom observations
   - Teachers were observed 3 times per year
   - 60 percent of overall evaluation score based on classroom observations

3. Rating based on classroom achievement level
   - Achievement of students on nationally normed subject assessment
   - 20 percent of overall evaluation score based on classroom achievement level

**District 13**

**Key program features**

- Teachers could receive a bonus for 1 overall performance measure that combined ratings based on school achievement growth and levels, classroom achievement growth, and classroom observations
- Set an absolute maximum bonus
- Teachers receiving a score of 4 on a 1 to 4 scale received a bonus

**Specific information on performance measures**

1. Rating based on school achievement growth and levels
   - Based on achievement growth and achievement levels on the state assessment
   - 20 percent of overall evaluation score based on school achievement growth and levels

2. Rating based on classroom achievement growth
   - For teachers in tested grades and subjects, based on value-added score on state assessment
   - For other teachers, based on student growth on student learning objectives
   - 20 percent of overall evaluation score based on classroom achievement growth

| **Cohort 2** |
|---|
| *District 13 (continued)*<br>3. Rating based on classroom observations<br>• Teachers were observed 3 times per year<br>• 60 percent of overall evaluation score based on classroom observations |

Source: District interviews from 2012 and 2013, grantees' Annual Performance Report (APR) documents, and technical assistance documents.

Note: Grantees submit an APR to the U.S. Department of Education that describes how educators are evaluated.

ELA = English language arts.

## Teachers

Table IV.4 in Chapter IV shows the percentage of the Cohort 1 districts in Years 1 and 2 that met the TIF grant goals for substantial, differentiated, and challenging to earn bonuses. Table D.10 compares the percentage of Cohort 1 districts that met these criteria to the percentage of both cohorts (Cohorts 1 and 2) that met these criteria in Year 1.

**Table D.10. Evaluation Districts Meeting TIF Grant Goals for Pay-for-Performance Bonuses for Teachers in Year 1, Cohort 1 and Cohorts 1 and 2 (Percentages)**

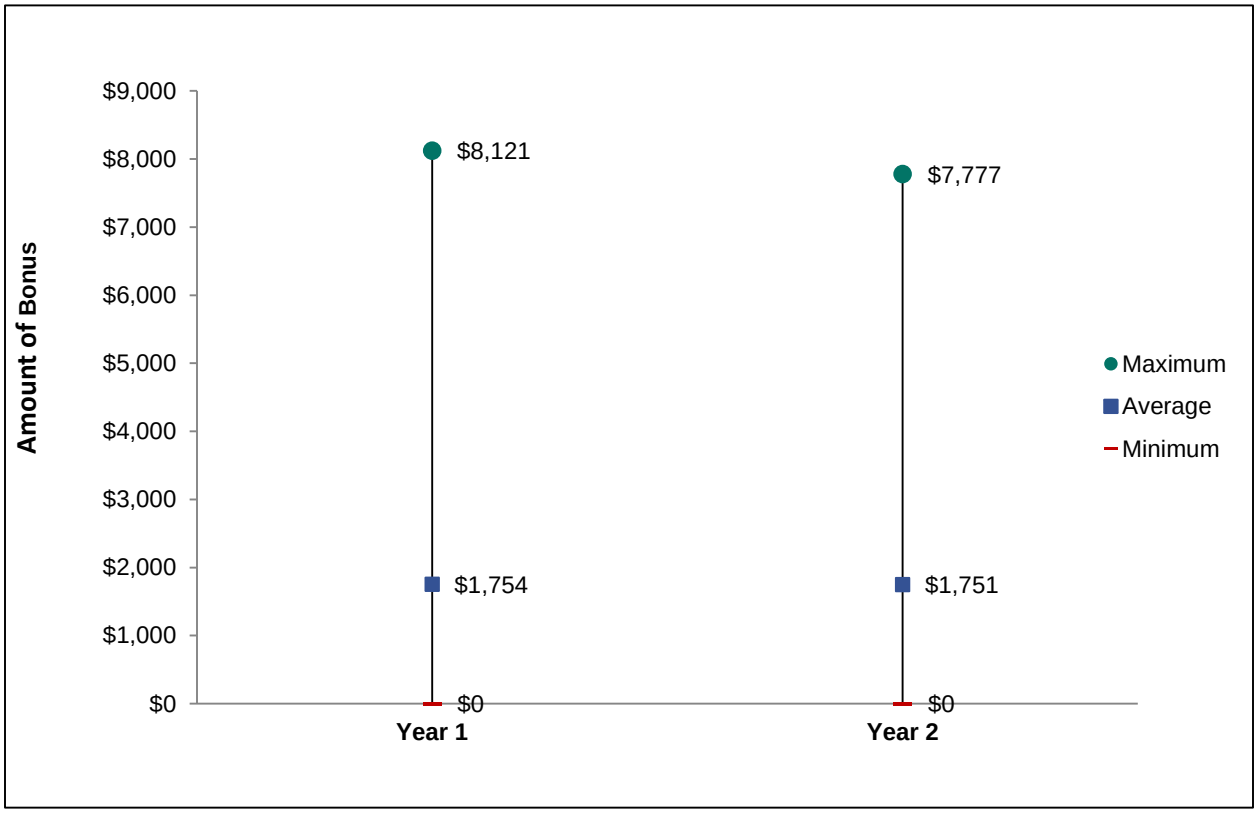
| TIF Grant Goal | Cohort 1 | Cohorts 1 and 2 |
|---|---|---|
| Substantial: Average bonus was at least 5 percent of average salary | 20 | 15 |
| Differentiated: Highest bonus was at least three times the average bonus | 70 | 69 |
| Challenging: Less than 50 percent of teachers received a pay-for-performance bonus | 20 | 31 |
| **Number of Districts** | **10** | **13** |

Source: Educator administrative data.

Figure IV.3 shows the minimum, average, and maximum pay-for-performance bonuses in Year 1 for teachers in Cohort 1, with each district equally weighted. By weighting each district equally, our findings in Chapter IV describe these bonuses for the average Cohort 1 district. Because our findings on educators' understanding and impact findings weight schools equally, Figure D.3 presents the minimum, average, and maximum pay-for-performance bonuses in Year 1 for teachers in Cohort 1, with districts weighted by the number of schools.

Figure D.4 compares the minimum, average, and maximum pay-for-performance bonuses in Year 1 for teachers in Cohort 1 to those for teachers in Cohorts 1 and 2. Like Figure IV.2, Figure D.4 weights each district equally.

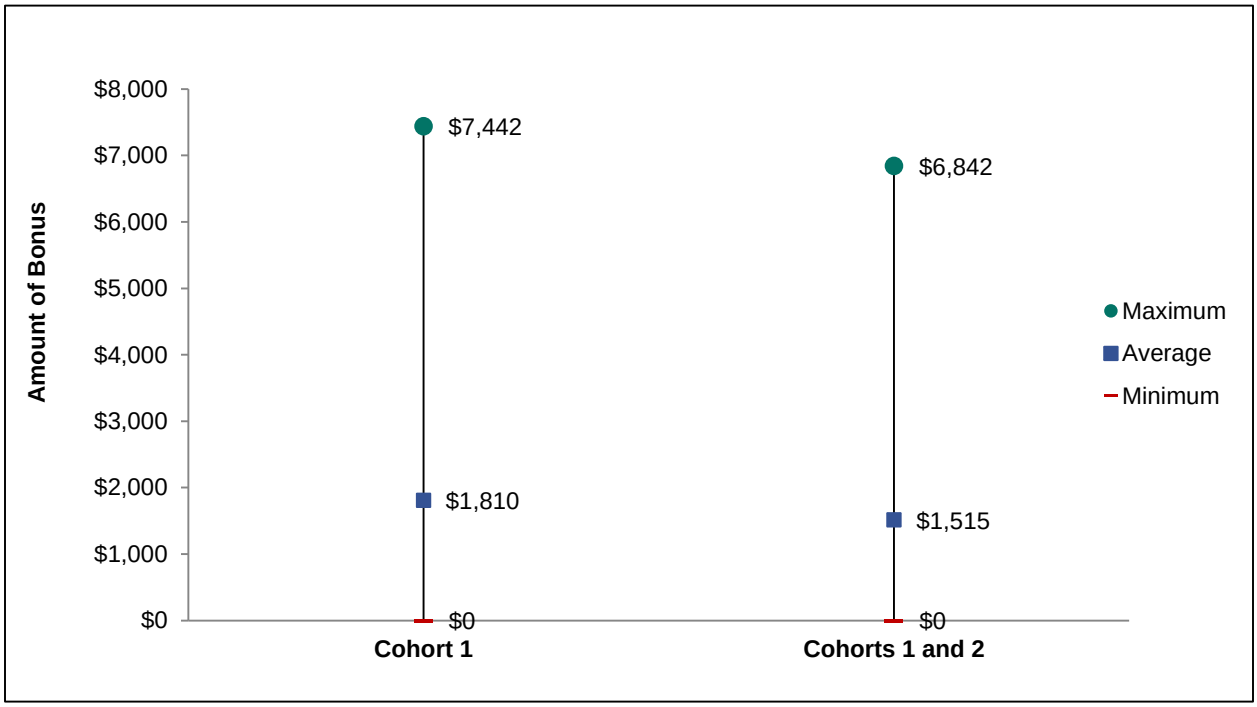
**Figure D.3. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers, with Districts Weighted by the Number of Schools, Cohort 1**




Source: Educator administrative data (N = 2,189 teachers in Year 1 and N = 2,207 teachers in Year 2).

Note: The statistics shown in the figure represent an equal-weighted average of the statistics from the schools in the Cohort 1 districts.

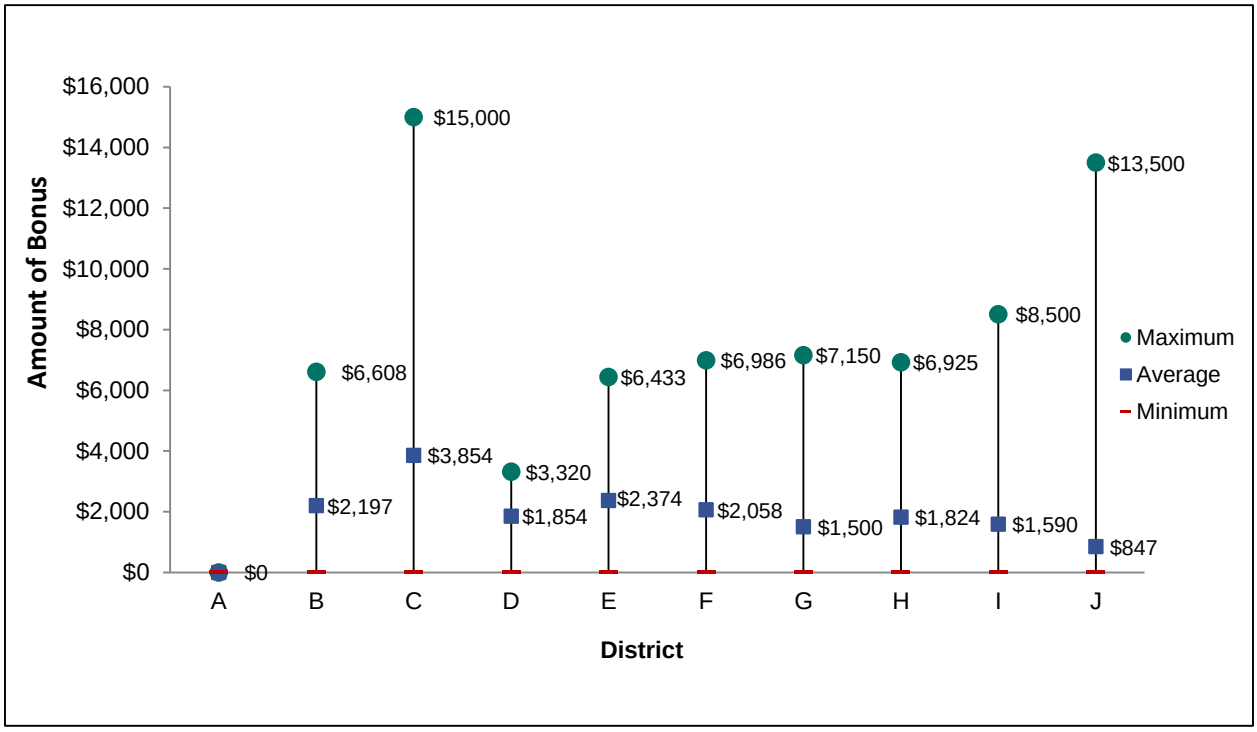
**Figure D.4. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers for Year 1, Cohort 1 and Cohorts 1 and 2**




Source: Educator administrative data (N = 2,189 teachers for Cohort 1 and N = 3,211 teachers for Cohorts 1 and 2).

Note: The statistics shown in the figure represent an equal-weighted average of the statistics from the 10 evaluation districts in Cohort 1.

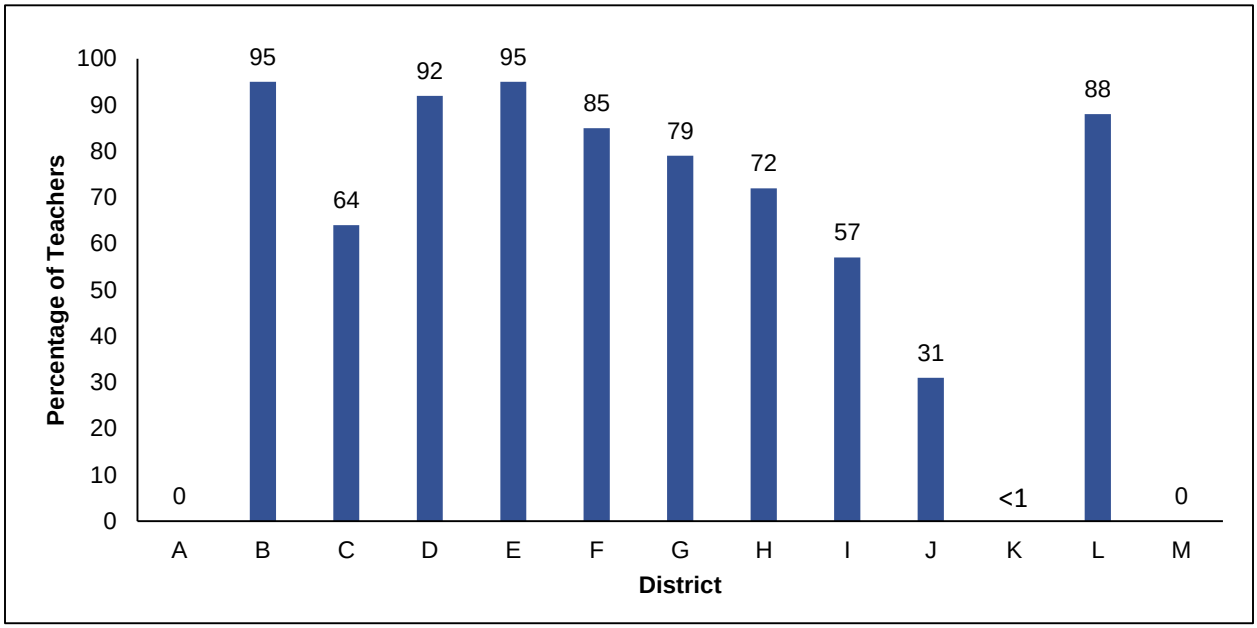
**Figure D.5. Distribution of Teachers' Pay-for-Performance Bonuses from TIF by District, Year 1, Cohort 1**




Source: Educator administrative data (N ranges from 73 teachers in District D to 432 in District J).

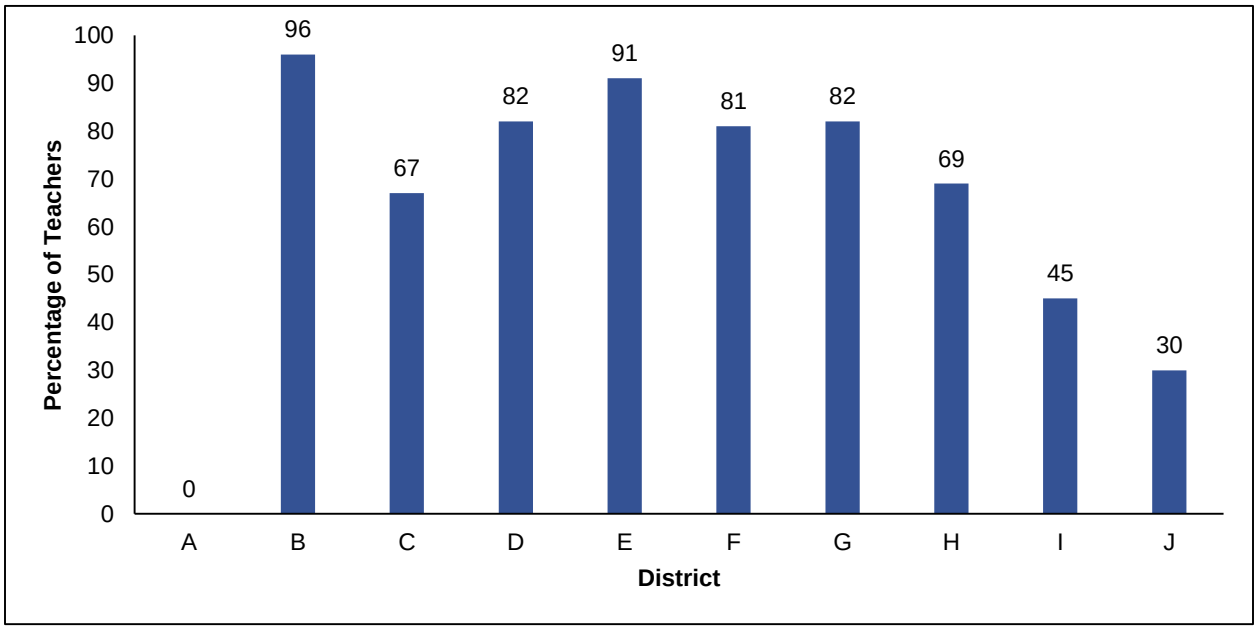
Applicants for the evaluation grants received guidance on the structure of their pay-for-performance bonus, including the example of *challenging* to earn bonuses, in which only those performing significantly better than the average (therefore, fewer than 50 percent) would receive a bonus. Figure IV.4 shows that, across districts, on average, more than 60 percent of treatment teachers received a bonus. Figure D.6 shows the percentage of teachers earning pay-for-performance bonuses in Year 1, by district, for Cohorts 1 and 2. Figure D.7 shows the percentage of teachers earning pay-for-performance bonuses in Year 2, by district, for Cohort 1.

**Figure D.6. Percentage of Teachers Earning Pay-for-Performance Bonuses in Year 1, by District, Cohorts 1 and 2**




Source: Educator administrative data (N ranges from 49 teachers in District L to 432 in District J).

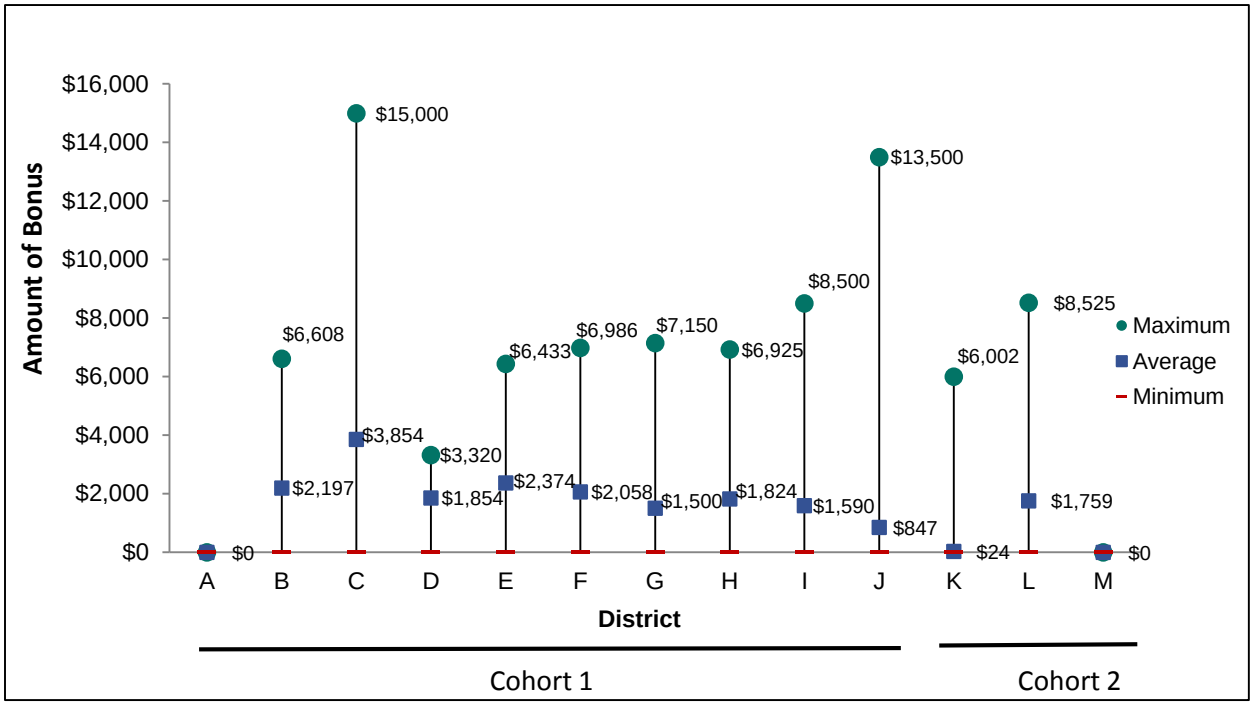
**Figure D.7. Percentage of Teachers Earning Pay-for-Performance Bonuses in Year 2, by District, Cohort 1**




Source: Educator administrative data (N ranges from 81 teachers in District E to 394 in District J).

In Chapter IV, we noted that the maximum bonus amounts for teachers varied substantially across districts. Figure IV.5 shows the distribution of pay-for-performance bonuses for teachers by district for Cohort 1 in Year 2. For comparison, we show the distribution of teachers' Year 1 pay-for-performance bonuses by district for Cohort 1 only (Figure D.5) and for Cohorts 1 and 2 (Figure D.8).

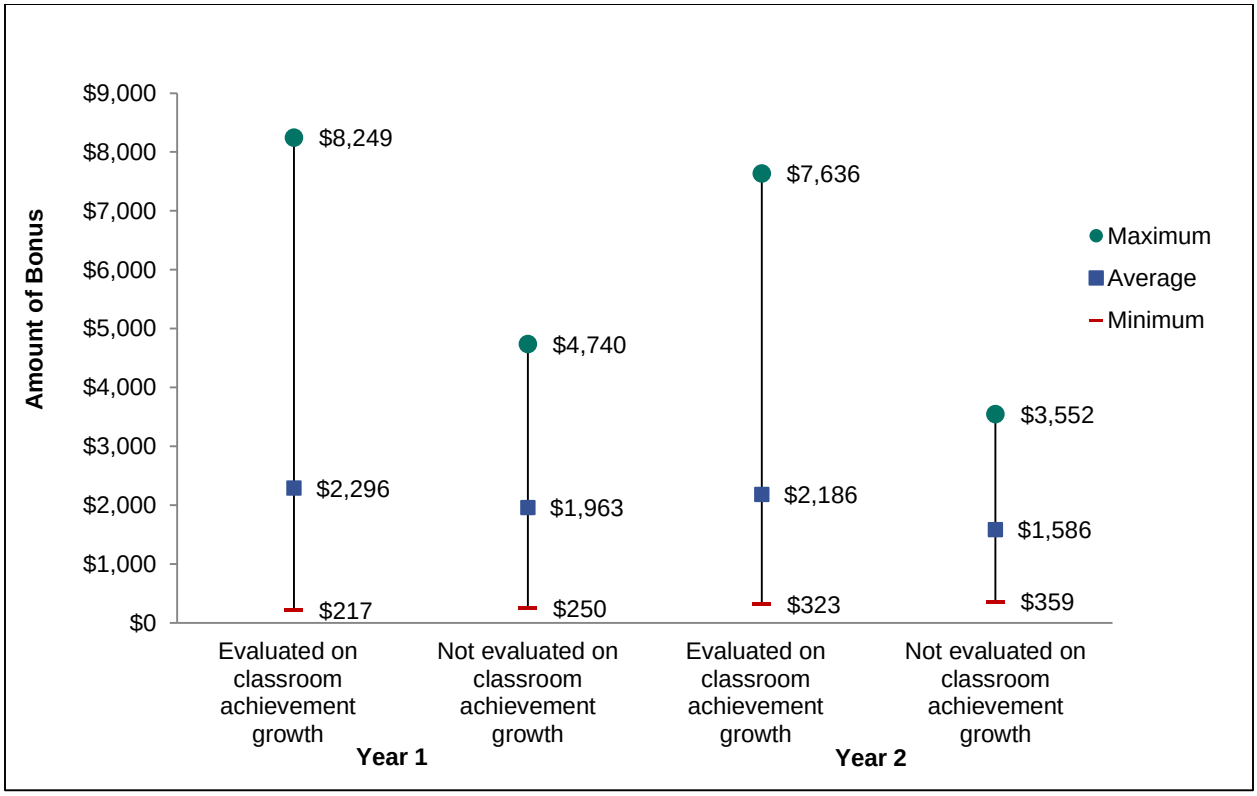
**Figure D.8. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers in Year 1 by District, Cohorts 1 and 2**




Source: Educator administrative data (N ranges from 49 teachers in District L to 432 in District J).

As discussed in Chapter IV, the six Cohort 1 districts that used classroom achievement growth measures linked a large share of the maximum performance bonus to those measures. Figure D.9 shows the minimum, average, and maximum pay-for-performance bonuses for teachers in the 6 of 10 Cohort 1 districts that evaluated teachers on classroom achievement growth for teachers who were and were not evaluated on their own students' achievement. As Figure D.9 shows, teachers who were evaluated on those measures generally could earn larger bonuses than those who were not. Teachers who were evaluated on classroom achievement growth in Year 2 earned maximum performance bonuses of at least $7,600, whereas teachers who were not evaluated on those measures earned maximum bonuses of less than $4,800.

**Figure D.9. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Teachers in Districts That Used Classroom Achievement Growth, Cohort 1**



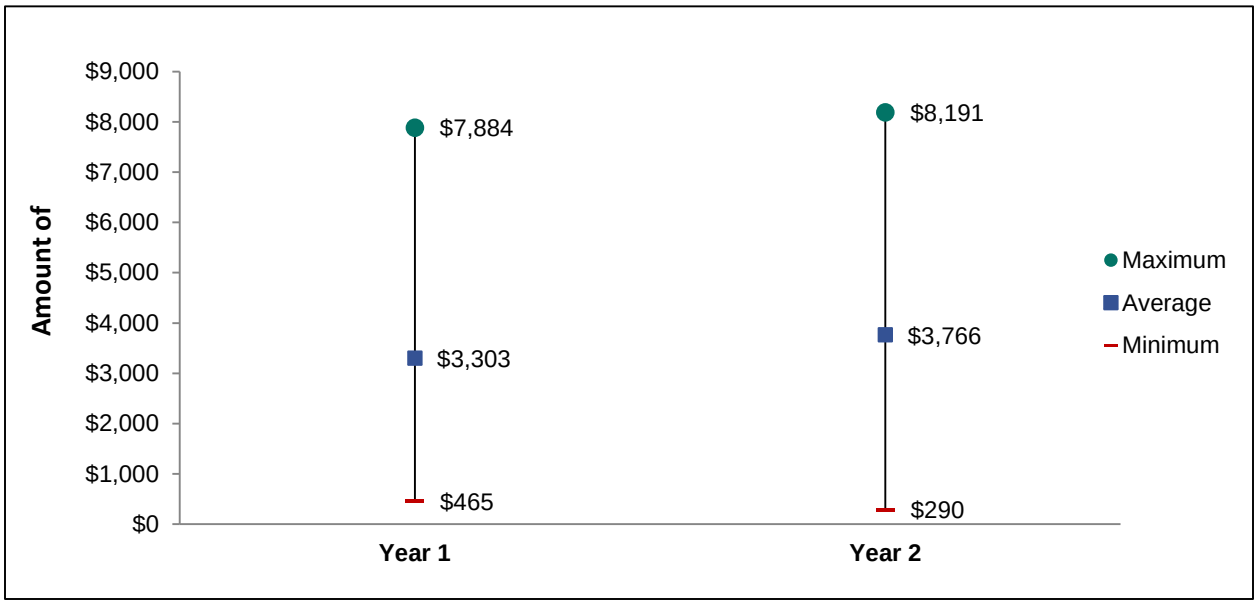

Source: Educator administrative data (537 teachers were evaluated on classroom achievement growth and 608 teachers were not evaluated on classroom achievement growth in Year 1; 642 teachers were evaluated on classroom achievement growth and 464 teachers were not evaluated on classroom achievement growth in Year 2). Six of 10 Cohort 1 districts evaluated teachers on classroom achievement growth.

## Principals

This section provides supplemental information on principals' pay-for-performance bonuses, similar to the previous section on teachers. Figure IV.7 shows the minimum, average, and maximum pay-for-performance bonuses in Years 1 and 2 for principals in Cohort 1, with each district equally weighted. Figure D.10 presents the minimum, average, and maximum pay-for-performance bonuses in Years 1 and 2 for principals in Cohort 1, with districts weighted by the number of schools.

Figure D.11 shows the same information as Figure IV.7 in Year 1 for Cohorts 1 and 2 combined. Similar to Figure IV.7, Figure D.11 findings give each district an equal weight.

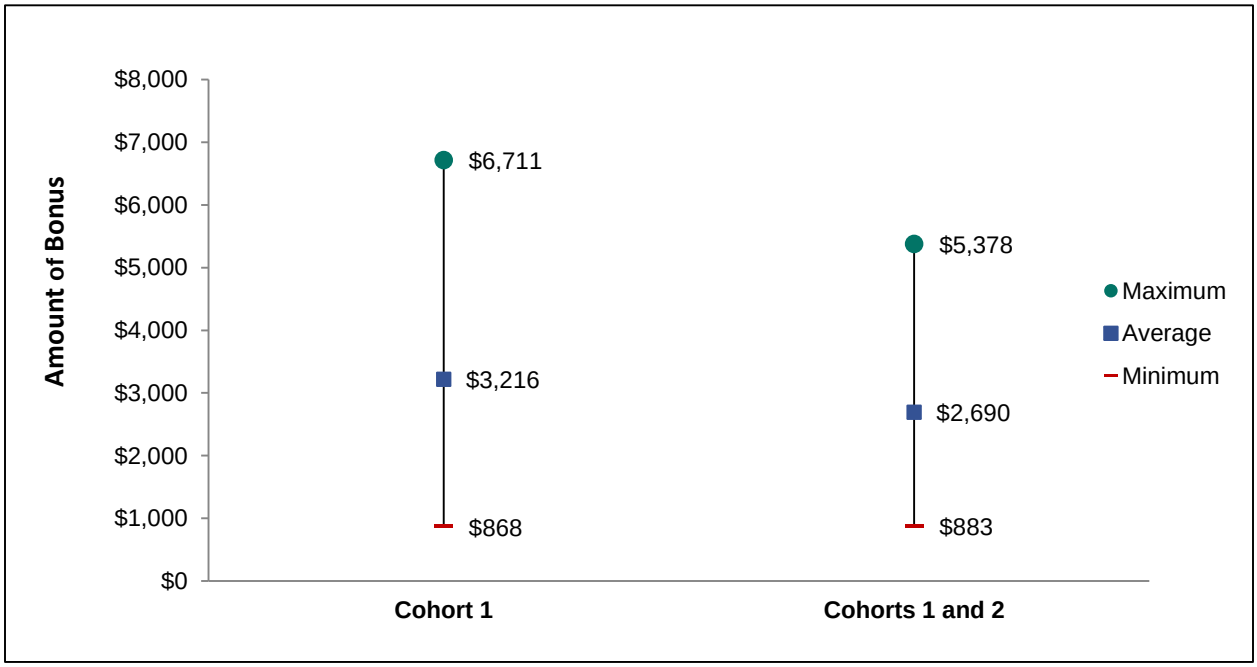
**Figure D.10. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Principals, with Districts Weighted by the Number of Schools, Cohort 1**




Source: Educator administrative data (N = 65 principals in Year 1 and N = 68 principals in Year 2).

Note: The statistics shown in the figure represent an equal-weighted average of the statistics from the schools in the Cohort 1 districts.

**Figure D.11. Minimum, Average, and Maximum Pay-for-Performance Bonuses for Principals for Year 1, Cohort 1 and Cohorts 1 and 2**

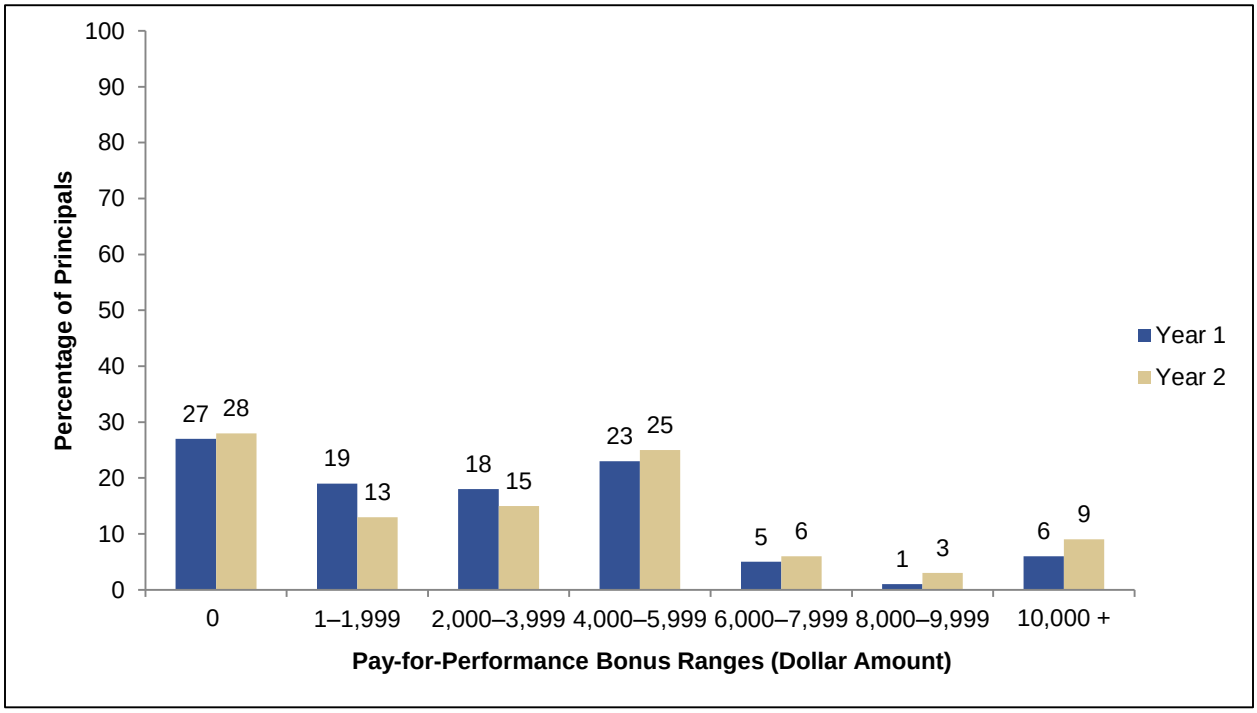



Source: Educator administrative data (N = 65 principals in Year 1, Cohort 1 and N = 91 principals in Year 1, Cohorts 1 and 2).

Note: The statistics shown in the figure represent an equal-weighted average of the statistics from the 10 evaluation districts in Cohort 1.

In Chapter IV, we noted that 20 percent of the districts met the guidance for challenging bonuses in Years 1 and 2 (Table IV.5). Figure D.12 illustrates the distribution of principals' pay-for-performance bonuses in Years 1 and 2 for Cohort 1. At least 70 percent of principals in each year received a bonus.

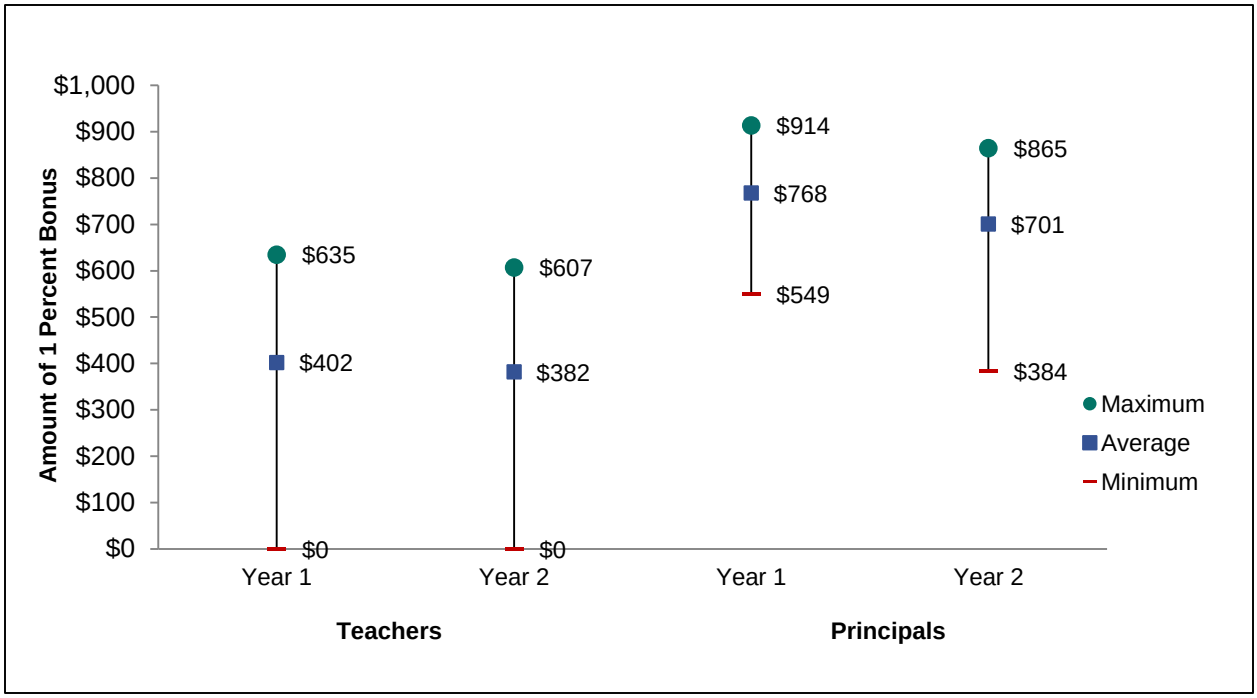
**Figure D.12. Distribution of Pay-for-Performance Bonuses for Principals, Cohort 1**




Source: Educator administrative data (N = 65 principals in Year 1 and N = 68 principals in Year 2).

Teachers and principals in control schools were expected to receive an automatic 1 percent bonus (see Chapter II). The 1 percent bonus ensured that all educators in evaluation schools received some benefit from participating in the study: either the opportunity to earn a pay-for-performance bonus or the automatic bonus. Figure D.13 presents the minimum, average, and maximum automatic 1 percent bonuses for Cohort 1 teachers and principals. As intended by the study design, the automatic 1 percent bonus provided to teachers and principals in control schools was small and did not vary substantially.

**Figure D.13. Minimum, Average, and Maximum Automatic 1 Percent Bonuses for Teachers and Principals, Cohort 1**




Source: Educator administrative data (Year 1: N = 2,157 teachers and N = 69 principals; Year 2: N = 2,259 teachers and N = 70 principals).

### Requirement 3: Additional Pay Opportunities

According to the study design, the only difference between treatment and control schools was the pay-for-performance bonus component of the TIF program. Educators in some schools (the treatment schools) were eligible for pay-for-performance, and educators in others (control schools) were not. As explained above, educators in control schools were expected to receive an automatic 1 percent bonus. All other aspects of the districts' TIF program (such as additional pay opportunities) should have been implemented the same in treatment and control schools.

Table D.11 shows the average and maximum payouts for additional pay and the percentage of teachers receiving additional pay for taking on extra roles across treatment and control schools for Cohort 1 in Years 1 and 2. Few teachers (less than 20 percent) received additional pay. Because most teachers received $0 in additional pay, the average amount teachers received (including those who received nothing) was notably less than the average pay-for-performance bonus that treatment teachers received ($1,760 in Year 2; Figure IV.3).

**Table D.11. Average and Maximum Amounts of Additional Pay Opportunities for Teachers, Cohort 1**

| Additional Pay Opportunities | Year 1 | Year 2 |
|---|---|---|
| Average amount for additional pay opportunities (dollars) | 452 | 502 |
| Maximum amount for additional pay opportunities (dollars) | 4,766 | 5,869 |
| Percentage of districts offering additional pay opportunities | 100 | 100 |
| Percentage of teachers receiving additional pay opportunities | 12 | 17 |
| **Number of Teachers** | **4,346** | **4,466** |

Source: Educator administrative data.

Table D.12 compares the amount of bonuses and additional pay received by Cohort 1 teachers in treatment and control schools in Years 1 and 2. As expected, the average bonus that treatment teachers received (a pay-for-performance bonus) was greater than the average bonus that control teachers received (a 1 percent bonus). In addition, as intended by the study design, the average amount of additional pay for extra roles or any other additional pay earned by teachers in treatment schools and control schools did not differ.

**Table D.12. Teacher Bonuses and Additional Pay, Cohort 1**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Pay-for-Performance (Treatment) or Automatic 1 Percent Bonuses (Control) | | | | | | |
| Average bonus (dollars) | 1,769 | 389 | 1,381* | 1,746 | 367 | 1,379* |
| Received bonus (percentage) | 59 | 80 | -21* | 56 | 78 | -22* |
| Roles and Responsibilities | | | | | | |
| Average additional pay (dollars) | 504 | 506 | -2 | 530 | 538 | -8 |
| Received pay (percentage) | 13 | 14 | -1* | 16 | 18 | -2* |
| Other Additional Pay[a] | | | | | | |
| Average additional pay (dollars) | 329 | 329 | 0 | 326 | 388 | -62* |
| Received pay (percentage) | 22 | 22 | 0 | 12 | 18 | -6* |
| Total Payouts[b] | | | | | | |
| Average payout (dollars) | 2,602 | 1,223 | 1,379* | 2,602 | 1,293 | 1,310* |
| Received payout (percentage) | 78 | 93 | -15* | 69 | 88 | -18* |
| **Number of Teachers** | **2,189** | **2,157** | | **2,207** | **2,259** | |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Other additional pay includes pay for factors such as working in a hard-to-staff school or subject area or professional development.

[b]Total payouts includes performance pay, automatic 1 percent bonuses, pay for additional roles and responsibilities, and other pay.

*Difference between treatment and control group is statistically significant at the 0.05 level, two-tailed test.

### Requirement 4: Professional Development

The TIF grant required that districts provide professional development linked to the measures of educator effectiveness. This support included professional development to help educators understand the measures being used to evaluate their performance, as well as feedback based on their actual performance ratings to help improve their instructional practices. Table D.13 shows that the percentage of teachers whom districts expected to receive the professional development required under the grant (the first two rows of Table D.13) did not differ substantially between Year 1 and Year 2. However, not all districts required teachers to participate in all professional development opportunities. Table D.14 shows the percentage of districts that reported if teachers had flexibility in choosing the professional development activities they attended.

**Table D.13. Percentages of Teachers Whom Districts Expected to Receive Professional Development Under TIF, Cohort 1**

| | Year 1 | Year 2 |
|---|---|---|
| Understanding performance measures of TIF program | 79 | 70 |
| Feedback based on TIF performance ratings | 53 | 58 |
| Understanding other components of TIF program | 90 | 68 |
| Other professional development topics | | |
| Differentiated instructional strategies based on student assessments | 25 | 53 |
| Instructional techniques and strategies | 67 | 61 |
| Aligning curricula to state or district standards | 57 | 43 |
| **Number of Districts** | **10** | **10** |

Source: District survey, 2012 and 2013.

**Table D.14. Teachers' Flexibility in Selecting Professional Development Opportunities, as Reported by Districts in Year 2, Cohort 1 (Percentages)**

| | Evaluation Districts |
|---|---|
| Flexible (e.g. selection of professional development made by individual teachers) | 20 |
| Semi-flexible (e.g. choice of professional development made by teachers in collaboration with other school or district staff) | 30 |
| Not flexible (e.g. professional development opportunities determined by someone other than teachers) | 30 |
| Unknown/ District did not provide information | 20 |
| **Number of Districts** | **10** |

Source: 2013 district interviews.

## Communication of TIF Program

We asked district administrators more detailed information on their communication activities during the district interviews. Table D.15 shows information from these interviews on who was responsible for communicating information about TIF to educators, whether districts adjusted their communication activities, the communication methods districts used, the frequency of the communication activities, and the topics of communication activities.

**Table D.15. Districts' Communication Activities in Year 2, Cohort 1 (Percentages)**

| | Evaluation Districts |
|---|---|
| Responsible for Majority of Communication about TIF | |
| District or grantee official | 70 |
| School-level staff (e.g. principal or lead teacher) | 10 |
| Unknown/ District did not provide information | 20 |
| Communication methods | |
| Written materials (including letters, email, brochures, program manuals, newsletters) | 60 |
| In person meeting (including staff meetings, small group presentations, orientations) | 90 |
| Conversations with stakeholders such as school boards, unions | 40 |
| District website | 30 |
| Media coverage (including social media) | 20 |
| Adjustment of Communication Approaches | |
| Mentioned during interview that they adjusted communication | 60 |
| Did not mention during interview that they adjusted communication | 40 |
| Frequency of communication | |
| Monthly | 40 |
| 2-3 times throughout the school year | 20 |
| Once during the school year | 10 |
| Don't know | 20 |
| Missing | 10 |
| Topics of communication | |
| Number, size, and distribution of bonus awarded for the 2011-2012 school year | 60 |
| Expectations about the number, size, and distribution of bonuses to be awarded based on 2012-2013 performance | 50 |
| Changes to the TIF program for the 2012-2013 school year | 30 |
| Don't know | 40 |
| **Number of Districts** | **10** |

Source: 2013 district interviews.

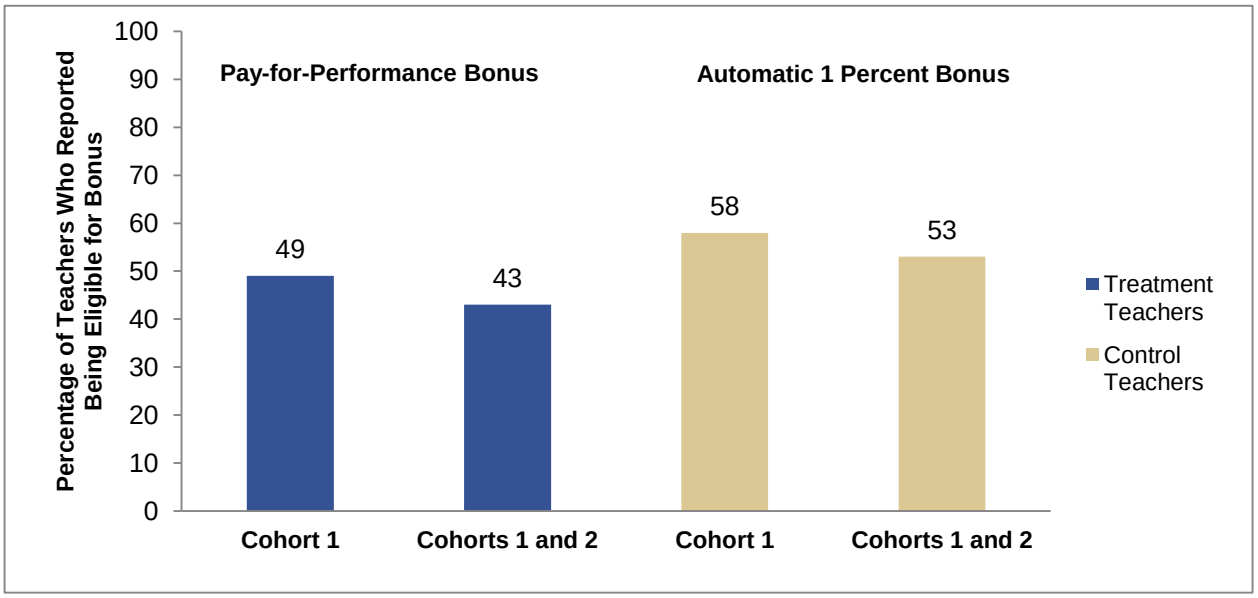
## Teacher and Principal Perspectives Regarding TIF Implementation

This section of the appendix provides additional details and supplemental analyses about educators' reported understanding of the TIF program.

### Educators' Understanding of their Eligibility for Pay-for-Performance Bonuses

Figures IV.8 and IV.9 show the percentages of Cohort 1 educators in treatment and control schools who reported they were eligible for either bonus in Years 1 and 2. Figure D.14 shows the percentage of teachers in treatment schools who reported they were eligible for a pay-for-performance bonus and the percentage of control teachers who reported they were eligible for an automatic 1 percent bonus in Year 1 for Cohort 1 compared to Cohorts 1 and 2 combined. Figure D.15 shows the same information for principals. When Year 1 analyses were based on Cohorts 1 and 2, similar, but somewhat smaller, percentages of teachers and principals reported being eligible for the correct type of bonus than Year 1 estimates based only on Cohort 1.

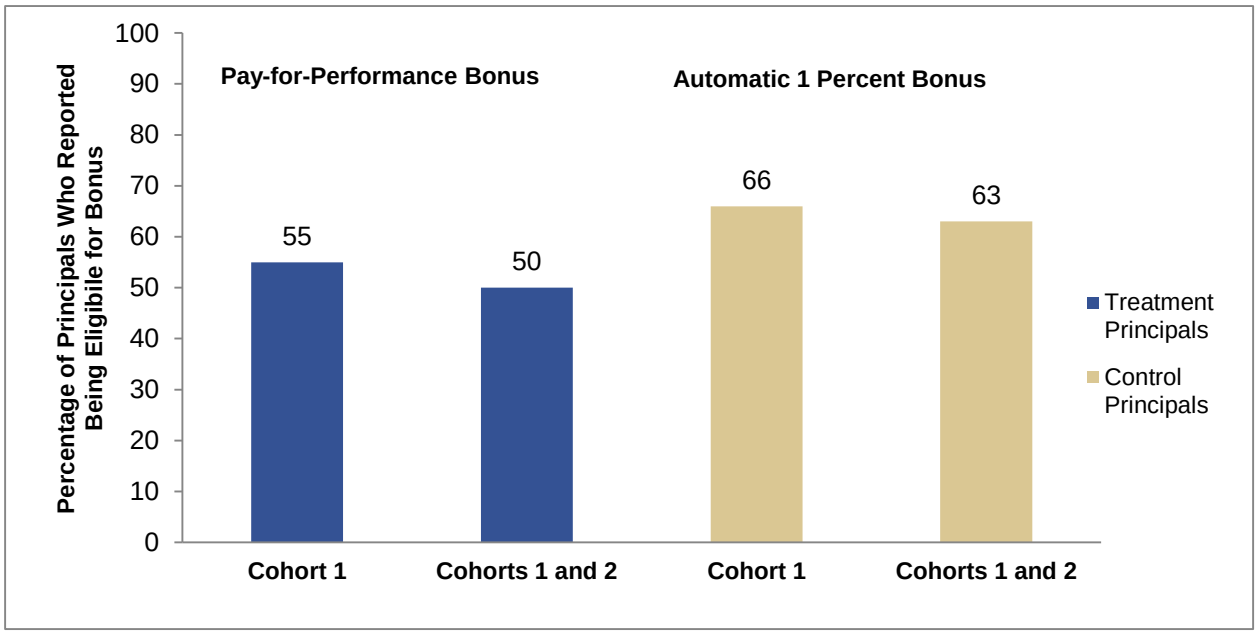
**Figure D.14. Teachers' Pay-for-Performance Bonus Eligibility in Year 1, as Reported by Teachers in Cohort 1 and Cohorts 1 and 2**




Source: Teacher surveys, 2012 and 2013.

Notes: A total of 377 treatment teachers in Cohort 1 and 520 in Cohorts 1 and 2 responded to the question about eligibility for a pay-for-performance bonus. A total of 381 control teachers in Cohort 1 and 497 in Cohorts 1 and 2 responded to the question about eligibility for an automatic 1 percent bonus.

**Figure D.15. Principals' Pay-for-Performance Bonus Eligibility in Year 1, as Reported by Principals in Cohort 1 and Cohorts 1 and 2**




Source: Principal surveys, 2012 and 2013.

Notes: A total of 64 treatment principals in Cohort 1 and 84 in Cohorts 1 and 2 responded to the question about eligibility for a pay-for-performance bonus. A total of 64 control principals in Cohort 1 and 86 in Cohorts 1 and 2 responded to the question about eligibility for an automatic 1 percent bonus.

Table D.16 shows the percentage of Cohort 1 educators who correctly reported their bonus eligibility as intended by the study design (also shown in Figures D.14 and D.15), but it also shows the percentage that misreported their eligibility. Specifically, it shows the percentage of educators in treatment schools who reported they were eligible for an automatic 1 percent bonus and the percentage of educators in control schools who reported they were eligible for a pay-for-performance bonus. Although more Cohort 1 educators correctly reported their eligibility in Year 2 than Year 1 (Figures IV.8 and IV.9), many educators continued to misreport their eligibility. For example, in Year 2, 38 percent of treatment teachers did not report being eligible for a pay-for-performance bonus, 40 percent of treatment teachers believed they were eligible for an automatic 1 percent bonus, and 17 percent of control teachers believed they were eligible for a pay-for-performance bonus.

**Table D.16. Bonus Eligibility as Reported by Teachers and Principals, Cohort 1**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| **Teachers** | | | | | | |
| Pay-for-performance | 49 | 17 | 32* | 62+ | 17 | 45* |
| Automatic 1 percent bonus | 39 | 58 | -19* | 40 | 80+ | -40* |
| **Number of Teachers—Range[a]** | **377-378** | **379-381** | | **448-458** | **456-461** | |
| **Principals** | | | | | | |
| Pay-for-performance | 55 | 13 | 42* | 90+ | 15 | 75* |
| Automatic 1 percent bonus | 27 | 66 | -39* | 31 | 85+ | -54* |
| **Number of Principals—Range[a]** | **63-64** | **63-64** | | **63-64** | **61** | |

Source: Teacher and principal surveys, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference between treatment and control group is statistically significant at the 0.05 level, two-tailed test.
+Difference between Year 1 and Year 2 within treatment status is statistically significant at the 0.05 level, two-tailed test.

As explained in Chapter IV, because understanding about eligibility for a bonus is critical for changing behavior, we explored how teacher understanding varied across districts, across schools within the same district, and within the same school. Table D.17 shows the percentage of the variation in teachers' understanding of their bonus eligibility that can be attributed to variation across districts, variation across schools within the same district, and variation across teachers within the same schools.[81] We found that most of the difference in teachers' understanding (more than 85 percent of the variation across treatment teachers and more than 74 percent of the variation across control teachers) occurs among teachers in the same school (Table D.17).

[81] We disaggregated the variance components by estimating a random effect model of bonus eligibility on intercepts for schools and districts that account for the nesting of teachers in schools and schools in districts. We estimated the model separately for treatment and control teachers.

**Table D.17. Percentages of Total Variance in Teachers' Understanding of Their Bonus Eligibility Attributable to Districts, Schools, and Teachers, Cohort 1**

| | Pay-for-Performance Bonus Eligibility (Treatment Schools) | | Automatic 1 Percent Bonus Eligibility (Control Schools) | |
|---|---|---|---|---|
| | Year 1 | Year 2 | Year 1 | Year 2 |
| Variation across districts | 12 | 5 | 11 | 16 |
| Variation across schools within districts | 3 | 3 | 4 | 10 |
| Variation across teachers within schools | 86 | 91 | 85 | 74 |
| **Number of Teachers** | **377** | **444** | **381** | **445** |
| **Number of Schools** | **66** | **66** | **66** | **66** |

Source: Teacher surveys, 2012 and 2013.

Tables D.18 and D.19 present subgroup results that examine district, principal, and teacher factors that might account for differences in treatment teachers' understanding of their eligibility for a performance bonus.

**Table D.18. Treatment Teachers' Reported Eligibility for Pay-for-Performance Bonuses in Year 2, by Districts' Characteristics, Cohort 1 (Percentages)**

| | Percentage of Teachers Reporting They Are Eligible for Pay-for-Performance Bonuses | Number of Treatment Teachers |
|---|---|---|
| **All Teachers (primary analysis)** | **62** | **444** |
| District additional compensation program available | | |
| (1) No additional compensation program available | 58 | 348 |
| (2) Additional compensation program available | 46 | 96 |
| Difference between (1) - (2) | 12 | |
| District adjusted aspects of communication based on lessons learned the prior year | | |
| (1) Adjusted communication | 73 | 274 |
| (2) Did not adjust communication | 57 | 170 |
| Difference between (1) - (2) | 16* | |
| District communication approach | | |
| (1) Centralized – relied primarily on district staff | 72 | 388 |
| (2) Decentralized – relied primarily on school staff | 64 | 56 |
| Difference between (1) - (2) | 8 | |
| District assessment of teachers' understanding of TIF | | |
| (1) Assessed understanding | 78 | 343 |
| (2) Did not assess understanding | 68 | 34 |
| Difference between (1) - (2) | 10 | |
| District uses Teacher Advancement Program (TAP) model | | |
| (1) TAP | 50 | 68 |
| (2) Non-TAP | 63 | 376 |
| Difference between (1) - (2) | -13 | |
| District expectations of teachers' participation in professional development for 2012–2013 | | |
| (1) At least 75 percent of teachers will participate in professional development | 63 | 228 |
| (2) Less than 75 percent of teachers will participate in professional development | 67 | 216 |
| Difference between (1) - (2) | -4 | |
| District Year 1 pay-for-performance bonus distribution method | | |
| (1) Pay-for-performance bonus paid in regular paycheck | 57 | 155 |
| (2) Pay-for-performance bonus paid in separate check | 56 | 68 |
| Difference between (1) - (2) | 1 | |
| District communication of actual bonuses | | |
| (1) Communicated the number, size, and distribution of bonus awarded for the 2011–2012 school year | 66 | 305 |
| (2) Did not communicate the number, size, and distribution of bonus awarded for the 2011–2012 school year | 56 | 139 |
| Difference between (1) - (2) | 11 | |
| District communication of expected bonuses | | |
| (1) communicated expectations about the number, size, and distribution of bonuses to be awarded based on 2012–2013 performance | 64 | 251 |
| (2) did not communicate expectations about the number, size, and distribution of bonuses to be awarded based on 2012–2013 performance | 59 | 193 |
| Difference between (1) - (2) | 4 | |

Source: Teacher and district surveys (2013) and district interviews (2013).

Notes: Subgroup means and hypothesis testing are based on a model with an indicator for the subgroup.

*Difference between subgroups is statistically significant at the 0.05 level, two-tailed test.

**Table D.19. Treatment Teachers' Reported Eligibility for Pay-for-Performance Bonuses in Year 2, by Principal Understanding and Teacher Characteristics, Cohort 1 (Percentages)**

| | Percentage of Teachers Reporting They Are Eligible for Pay-for-Performance Bonuses | Number of Treatment Teachers |
|---|---|---|
| **All Teachers (primary analysis)** | **62** | **444** |
| **Subgroup Analysis By Principal Understanding** | | |
| Principal understanding of teachers' eligibility | | |
| (1) Principal correctly reported teachers' eligibility | 57 | 222 |
| (2) Principal incorrectly reported teachers' eligibility | 60 | 171 |
| Difference between (1) - (2) | -3 | |
| **Subgroup Analyses By Teacher Characteristics** | | |
| Teaching assignment | | |
| (1) Tested | 68 | 232 |
| (2) Untested | 54 | 212 |
| Difference between (1) - (2) | 14 | |
| Pay-for-performance bonus receipt in Year 1 | | |
| (1) Received pay-for-performance bonus in Year 1 | 64 | 279 |
| (2) Did not received pay-for-performance bonus in Year 1 | 49 | 165 |
| Difference between (1) - (2) | 15 | |
| Teacher participated in professional development about TIF performance measures | | |
| (1) Teacher participated in professional development | 58 | 277 |
| (2) Teacher did not participate in professional development | 59 | 156 |
| Difference between (1) - (2) | -1 | |
| Mentoring role | | |
| (1) Teacher is a mentor teacher | 57 | 115 |
| (2) Teacher is not a mentor teacher | 63 | 324 |
| Difference between (1) - (2) | -5 | |
| (1) Teacher has a mentor teacher | 58 | 222 |
| (2) Teacher does not have a mentor teacher | 63 | 219 |
| Difference between (1) - (2) | -5 | |

Source: Teacher, principal, and district surveys (2013) and district interviews (2013).

Notes: Subgroup means and hypothesis testing are based on a model with an indicator for the subgroup. None of the differences between subgroups are significantly different.

## Educators' Understanding of the Potential Amounts of Pay-for-Performance Bonuses

Figures IV.11 and IV.12 show the actual and reported maximum pay-for-performance bonuses for teachers and for principals, respectively, for Cohort 1 in Years 1 and 2. For teachers and principals who reported being eligible for the bonus but left the amount missing, bonus amounts were imputed through multiple imputation methods (see Appendix B). Teachers' and principals' amounts are based on survey responses, with each school receiving an equal weight. Districts' expected and actual maximum bonus amounts are based on district survey responses and administrative data, with each district receiving an equal weight. This section shows analyses that do not use imputed values for missing data, and analyses that calculate districts' reported and actual maximum bonus amounts weighting each school equally.

Table D.20 shows the maximum possible bonus amounts as reported by educators with (1) missing values imputed (as shown in Figures IV.11 and IV.12), and (2) non-imputed bonus amounts. Table D.20 shows that our results are similar if we do not impute the missing bonus amounts.

**Table D.20. Educators' Reports on the Maximum Possible Bonus Amount: Imputed and Non-Imputed Bonus Amounts, Cohort 1**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| **Teachers** | | | | | | |
| Pay-for-performance | | | | | | |
| Imputed | 3,026 | 388 | 2,638* | 2,876 | 501 | 2,375* |
| Non-Imputed | 2,804 | 293 | 2,512* | 2,823 | 460 | 2,363* |
| Automatic 1 percent bonus | | | | | | |
| Imputed | 815 | 1,076 | -261 | 988 | 952 | 37 |
| Non-Imputed | 578 | 970 | -392 | 747 | 764 | -17 |
| **Number of Teachers—Range[a]** | **196-224** | **190-222** | | **194-232** | **185-252** | |
| **Principals** | | | | | | |
| Pay-for-performance | | | | | | |
| Imputed | 4,743 | 520 | 4,223* | 6,097 | 321 | 5,776* |
| Non-Imputed | 4,317 | 207 | 4,110* | 5,960 | 321 | 5,639* |
| Automatic 1 percent bonus | | | | | | |
| Imputed | 1,927 | 1,081 | 846 | 1,132 | 1,253 | -121 |
| Non-Imputed | 1,749 | 979 | 770 | 849 | 992 | -143 |
| **Number of Principals—Range[a]** | **56-64** | **58-64** | | **60-64** | **46-61** | |

Source: Teacher and principal surveys (2012 and 2013).

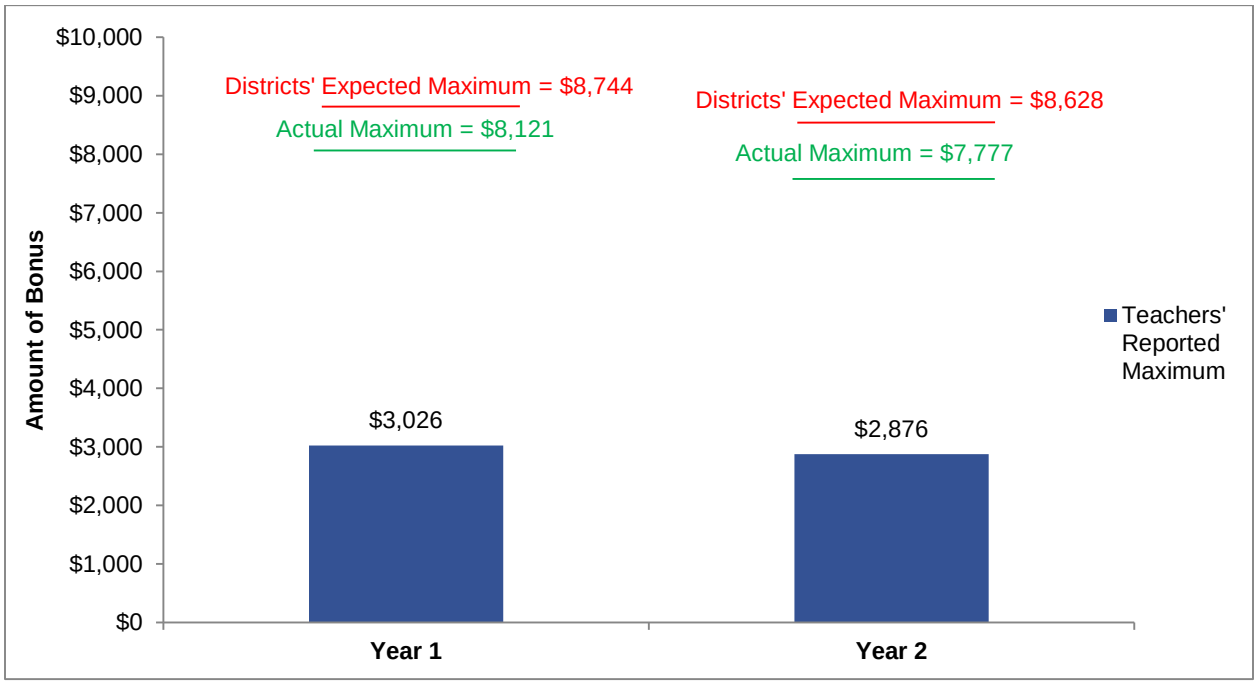
Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference between treatment and control group is statistically significant at the 0.05 level, two-tailed test.

Figures D.16 and D.17 show the actual and reported maximum pay-for-performance bonuses for teachers and for principals with the districts weighted by the number of schools. Unlike Figures IV.11 and IV.12, Figures D.16 and D.17 compare districts' amounts to educators' reported amounts using the same weighting approach. These figures show that our results are similar if we only use school weights.

**Figure D.16. Actual and Reported Maximum Pay-for-Performance Bonus for Teachers in Treatment Schools, with Districts Weighted by the Number of Schools, Cohort 1**

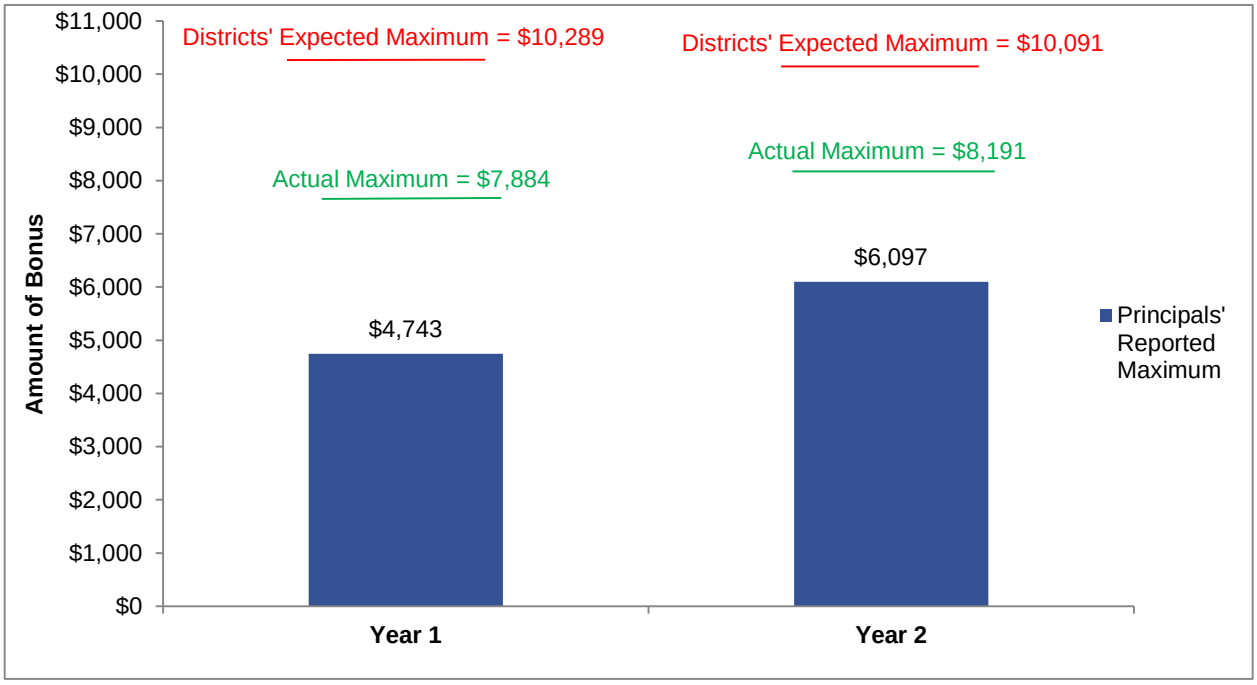



Source: Teacher survey (2012 and 2013), district interviews, and administrative data.

Notes: Teachers' reports are based on data for teachers in tested grades and subjects. Districts' reports and payouts are based on data for all teachers. All estimates were calculated weighting schools equally.

A total of 196 treatment teachers and 214 control teachers in tested grades and subjects responded to this survey question in Year 1. A total of 218 treatment teachers and 246 control teachers in tested grades and subjects responded to this survey question in Year 2. The maximum bonus amount was set to zero for all respondents who indicated they were ineligible for a bonus. For teachers who reported being eligible for the bonus but left the amount missing, bonus amounts were imputed through multiple imputation methods. This led to 27 additional responses for treatment teachers and 7 for control teachers in Year 1 and to 14 additional responses for treatment teachers and 6 for control teachers in Year 2. See Appendix B for additional discussion on the imputation methods. Appendix D, Table D.20 shows that our results are similar if we do not impute the missing bonus amounts.

**Figure D.17. Actual and Reported Maximum Pay-for-Performance Bonus for Principals in Treatment Schools, with Districts Weighted by the Number of Schools, Cohort 1**




Source: Principal survey (2012 and 2013), district interviews, and administrative data.

Notes: All estimates were calculated weighting schools equally. A total of 56 treatment principals and 60 control principals responded to this survey question in Year 1. A total of 61 treatment principals and 61 control principals responded to this survey question in Year 2. The maximum bonus amount was set to zero for all respondents who indicated they were ineligible for a bonus. For educators who reported being eligible for the bonus but left the amount missing, bonus amounts were imputed through multiple imputation methods. This led to 8 additional responses for treatment teachers and 3 for control teachers in Year 1 and to 2 additional responses for treatment teachers and 0 for control teachers in Year 2. See Appendix B for additional discussion on the imputation methods. Appendix D, Table D.20 shows that our results are similar if we do not impute the missing bonus amounts.

## Educators' Understanding of and Experiences with Professional Development

The TIF grant required that teachers receive professional development focused on understanding performance measures used in TIF and feedback based on their performance ratings. This requirement applied equally to teachers in treatment and control schools. Tables D.21 and D.22 show that teachers in treatment and control schools reported similar professional development experiences. These tables also support the finding discussed in Chapter IV that more than half of teachers reported they received the professional development required under the TIF grant but indicated they received only a few hours.

**Table D.21. Professional Development Teachers Reported Receiving or Expecting to Receive During the 2012–2013 School Year (Year 2), Cohort 1 (Percentages)**

| Professional Development Topics | Treatment | Control | Difference |
|---|---|---|---|
| Understanding components of TIF | 65 | 70 | -4 |
| Understanding performance measures of TIF | 63 | 68 | -6* |
| Feedback based on TIF performance ratings | 52 | 57 | -5* |
| Differentiated instructional strategies based on student assessments | 75 | 76 | 0 |
| Instructional techniques and strategies | 87 | 90 | -3 |
| Aligning curricula to state or district standards | 81 | 85 | -4 |
| **Number of Teachers—Range[a]** | **435-437** | **438-440** | |

Source: Teacher survey, 2013.

Note: The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Table D.22. Hours of Expected Professional Development for the 2012–2013 School Year, as Reported by Teachers (Year 2), Cohort 1 (Averages)**

| Professional Development Topics | Treatment | Control | Difference |
|---|---|---|---|
| Understanding components of TIF | 4 | 5 | -1 |
| Understanding performance measures of TIF | 3 | 4 | -1* |
| Feedback based on TIF performance ratings | 3 | 4 | -1 |
| Differentiated instructional strategies based on student assessments | 8 | 7 | 1 |
| Instructional techniques and strategies | 15 | 12 | 3* |
| Aligning curricula to state or district standards | 9 | 8 | 1* |
| **Number of Teachers—Range[a]** | **219-365** | **245-380** | |

Source: Teacher survey, 2013.

Notes: Table reports hours of professional development among teachers who indicated they had received or were planning to receive that type of professional development during the 2012–2013 school year. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the 0.05 level, two-tailed test.

THIS PAGE IS INTENTIONALLY BLANK

# APPENDIX E

# SUPPLEMENTAL FINDINGS ON IMPACTS OF PAY-FOR-PERFORMANCE ON EDUCATORS' ATTITUDES AND BEHAVIORS FOR CHAPTER V

THIS PAGE IS INTENTIONALLY BLANK

This appendix supplements the findings presented in Chapter V. As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts whose schools were randomly assigned in spring and summer 2011 were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. At the time of this report, Cohort 1 had completed two years of implementation, 2011–2012 and 2012–2013, referred to as Years 1 and 2. Cohort 2 districts had completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort.

Tables E.1 through E.7 present impact estimates for the first year of TIF implementation using all evaluation schools (Cohorts 1 and 2) and additional findings based on teachers' subgroups for Cohort 1 only. Tables E.8 through E.10 provide evidence on the impact of pay-for-performance on additional measures for Cohort 1: principals' hiring autonomy, staffing, and compensation decisions. Although these factors are not the main drivers of teachers' productivity or mobility captured in our logic model, they may still contribute to teachers' school environment and job satisfaction.

## Year 1 Impacts for Cohort 1 Schools Compared to Cohorts 1 and 2

In Chapter V, we presented impact estimates based on Cohort 1 schools that have implemented the program for two full years. Here, we show estimates for the first year of implementation (Year1) for all study schools that have implemented the program, combining Cohorts 1 and 2. These tables also include the Year 1 estimates for Cohort 1 only for easy comparison with the Year 1 estimates based on both cohorts.

**Table E.1. Teachers' Satisfaction with Professional Opportunities, Evaluation System, and School Environment, Cohorts 1 and 2 (Percentages Who Are "Somewhat" or "Very" Satisfied)**

| | Year 1 (Cohort 1) | | | Year 1 (Cohorts 1 and 2) | | |
|---|---|---|---|---|---|---|
| Satisfaction Dimension | Treatment | Control | Impact | Treatment | Control | Impact |
| Opportunities for Pay and Development | | | | | | |
| Opportunities for professional advancement | 67 | 75 | -9* | 67 | 73 | -6* |
| Opportunities to enhance skills | 76 | 78 | -2 | 76 | 77 | -1 |
| Opportunities to earn extra pay | 62 | 57 | 6 | 66 | 63 | 3 |
| Evaluation System | | | | | | |
| Use of student achievement scores to assess performance | 66 | 67 | -1 | 58 | 63 | -6* |
| School Environment | | | | | | |
| Recognition of accomplishments | 54 | 61 | -7* | 53 | 58 | -6* |
| Quality of interaction with colleagues | 75 | 81 | -6* | 75 | 82 | -7* |
| Colleagues' efforts | 83 | 85 | -1 | 81 | 83 | -2 |
| School morale | 50 | 55 | -5 | 45 | 52 | -7 |
| Job Satisfaction | | | | | | |
| Overall job satisfaction | 68 | 73 | -5 | 65 | 69 | -4 |
| **Number of Teachers—Range[a]** | **387-391** | **392-399** | | **538-543** | **515-524** | |

Source: Teacher survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table E.2. Principals' Satisfaction with Professional Opportunities, Evaluation System, and School Environment, Cohorts 1 and 2 (Percentages Who Are "Somewhat" or "Very" Satisfied)**

| | Year 1 (Cohort 1) | | | Year 1 (Cohorts 1 and 2) | | |
|---|---|---|---|---|---|---|
| Satisfaction Dimension | Treatment | Control | Impact | Treatment | Control | Impact |
| Opportunities for Pay and Development | | | | | | |
| Opportunities to enhance skills | 92 | 95 | -3 | 94 | 94 | 1 |
| Opportunities to earn extra pay | 72 | 66 | 6 | 70 | 62 | 8 |
| Evaluation System | | | | | | |
| Feedback on my performance | 84 | 87 | -3 | 81 | 85 | -4 |
| School Environment | | | | | | |
| Recognition of accomplishments | 78 | 82 | -4 | 76 | 78 | -2 |
| Quality of interaction with colleagues | 90 | 97 | -7 | 92 | 95 | -3 |
| Colleagues' efforts | 93 | 98 | -5 | 91 | 98 | -7 |
| School morale | 71 | 87 | -16* | 75 | 84 | -9 |
| **Number of Principals—Range[a]** | **63-64** | **59-61** | | **84-85** | **79-83** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table E.3. Teachers' Attitudes Toward TIF Program, Cohorts 1 and 2 (Percentages Who "Agree" or "Strongly Agree")**

| | Year 1 (Cohort 1) | | | Year 1 (Cohorts 1 and 2) | | |
|---|---|---|---|---|---|---|
| Statement | Treatment | Control | Impact | Treatment | Control | Impact |
| Teachers who do the same job should receive the same pay | 57 | 58 | -1 | 58 | 60 | -2 |
| Standardized student test scores in my district measure what students have learned | 35 | 33 | 2 | 33 | 30 | 4 |
| My principal is a good judge of teacher talent | 67 | 73 | -6 | 67 | 74 | -8* |
| I am glad that I am participating in the TIF program | 66 | 65 | 1 | 69 | 64 | 4 |
| My job satisfaction has increased due to the TIF program | 28 | 33 | -5 | 31 | 34 | -3 |
| I feel increased pressure to perform due to the TIF program | 65 | 53 | 11* | 63 | 51 | 13* |
| I have less freedom to teach the way I would like to teach due to the TIF program | 34 | 35 | 0 | 36 | 33 | 2 |
| The TIF program has harmed the collaborative nature of teaching | 23 | 24 | -1 | 27 | 25 | 1 |
| The TIF program has caused teachers to work more effectively | 49 | 46 | 3 | 46 | 45 | 0 |
| The TIF program is fair | 53 | 58 | -5 | 54 | 58 | -3 |
| The process used to determine how bonuses are determined was adequately explained to me | 67 | 59 | 8* | 60 | 54 | 7* |
| **Number of Teachers—Range[a]** | **381-388** | **382-398** | | **491-535** | **474-520** | |

Source: Teacher survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table E.4. Principals' Attitudes Toward TIF Program, Cohorts 1 and 2 (Percentage Who "Agree" or "Strongly Agree")**

| | Year 1 (Cohort 1) | | | Year 1 (Cohorts 1 and 2) | | |
|---|---|---|---|---|---|---|
| Statement | Treatment | Control | Impact | Treatment | Control | Impact |
| The TIF program has been clearly communicated to me | 83 | 89 | -6 | 80 | 85 | -5 |
| This school has less chance of earning a bonus because of the characteristics of our student population | 22 | 20 | 3 | 22 | 23 | -1 |
| The evaluation system omits important aspects of school administration that should be considered | 30 | 30 | 0 | 41 | 38 | 4 |
| The TIF program contributes to greater collegiality and professionalism among the staff at this school | 49 | 55 | -6 | 45 | 55 | -9 |
| Teachers at this school are more comfortable with frequent formal observations of their teaching because of the TIF program | 54 | 63 | -9 | 51 | 56 | -6 |
| Parents and the school community believe the TIF program is important | 39 | 48 | -8 | 38 | 42 | -4 |
| The TIF program is likely to continue for the foreseeable future | 85 | 87 | -2 | 79 | 83 | -4 |
| I played an important role in implementing the TIF program at my school | 82 | 84 | -2 | 78 | 79 | -1 |
| **Number of Principals—Range[a]** | **62-65** | **60-64** | | **83-86** | **80-86** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

## Additional Findings on Teachers' Attitudes by Subgroup

Tables E.5 through E.7 show supplementary analysis of teachers' satisfaction and attitudes in the second year of TIF implementation. Table E.5 shows the impacts of pay-for-performance on teachers' satisfaction with their professional opportunities, evaluation system, and school environment by subgroups based on teaching assignment and teaching experience. Table E.6 examines treatment teachers' satisfaction on these dimensions by whether the teacher received a bonus based on their Year 1 performance. Table E.7 shows the impacts of pay-for-performance on teachers' attitudes toward their job and the TIF program by subgroups based on teaching assignment and teaching experience.

**Table E.5. Impacts of Pay-for-Performance on Teacher Satisfaction Measures for Teacher Subgroups, Year 2, Cohort 1 (Percentage Points)**

| | Impacts on Whether Teachers Were "Somewhat" or "Very" Satisfied with… | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subgroup | Feedback on My Performance | Use of Student Achievement Scores to Measure Performance | Opportunities for Professional Advancement | Opportunities to Enhance My Skills | Opportunities to Earn Extra Pay | Recognition of Accomplishments | Quality of Interactions with Colleagues | Colleagues' Efforts | School Morale | Overall Job Satisfaction | **Number of Teachers** |
| All Teachers (primary analysis) | -5* | -9* | -3 | -1 | 9* | -6* | 0 | 0 | -1 | -1 | **892-896** |
| Teaching Assignment | | | | | | | | | | | |
| (1) Tested grades and subjects | -6 | -10* | -7 | -2 | 8 | -7 | -3 | -1 | -4 | -6 | **484-487** |
| (2) Nontested grades and subjects | -3 | -7 | 2 | 0 | 10 | -6 | 3 | 2 | 3 | 5 | **407-410** |
| **Difference between subgroup (1) - (2)** | **-3** | **-4** | **-9** | **-1** | **-2** | **-2** | **-6** | **-3** | **-7** | **-11** | |
| Teacher Experience | | | | | | | | | | | |
| (1) Less than 5 years | 5 | 5 | 8 | 3 | 9 | 5 | -3 | -10 | 14* | 10 | **191-193** |
| (2) 5 to 15 years | -3 | -9* | -3 | -2 | 7 | -3 | 0 | 3 | 0 | 0 | **453-454** |
| (3) Greater than 15 years | -15* | -16* | -10 | -2 | 10 | -21* | 2 | 2 | -12 | -10 | **247-250** |
| **Difference between subgroups (1) - (2)** | **8** | **14** | **11** | **4** | **2** | **7** | **-2** | **-13** | **14** | **10** | |
| **Difference between subgroups (3) - (2)** | **-12** | **-7** | **-8** | **0.0** | **3** | **-18** | **3** | **-1** | **-12** | **-9** | |

Source: Teacher survey, 2013.

Note: The difference between the treatment group and the control group is adjusted for block fixed effects and school-level characteristics at time of randomization. All teacher estimates come from Table V.1. Subgroup-specific impact estimates and hypothesis tests are based on a model with a treatment dummy and interaction(s) between the treatment and the subgroup(s) along with main effect for subgroup(s) using the pooled sample.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table E.6. Treatment Teachers' Satisfaction by Bonus Receipt, Year 2, Cohort 1 (Percentages who "Agree" or "Strongly Agree")**

| | Treatment Teachers | | |
|---|---|---|---|
| Statement | Received a Bonus After Year 1 | Did Not Receive a Bonus After Year 1 | Difference |
| Opportunities for Pay and Development | | | |
| Opportunities for professional advancement | 78 | 72 | 6 |
| Opportunities to enhance skills | 84 | 82 | 2 |
| Opportunities to earn extra pay | 61 | 66 | -5 |
| Evaluation System | | | |
| Use of student achievement scores to assess teacher effectiveness | 54 | 62 | -8 |
| Feedback on teacher performance | 78 | 73 | 5 |
| School Environment | | | |
| Recognition of accomplishments | 60 | 52 | 7 |
| Quality of interaction with colleagues | 82 | 77 | 6 |
| Colleagues' efforts | 90 | 78 | 12 |
| School morale | 52 | 47 | 4 |
| Job Satisfaction | | | |
| Overall job satisfaction | 73 | 64 | 9 |
| **Number of Teachers—Range[a]** | **279-282** | **163-166** | |

Source: Teacher survey (2013) and educator administrative data.

Notes: None of the differences are statistically significant at the .05 level. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

**Table E.7. Impacts of Pay-for-Performance on Teacher Attitude Measures for Teacher Subgroups, Year 2, Cohort 1 (Percentage Points)**

| | Impacts on Whether Teachers Responded They "Agreed" or "Strongly Agreed" with… | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subgroup | Teachers Who Do the Same Job Should Receive the Same Pay | Standardized Student Test Scores in My District Measure What Students Have Learned | My Principal Is a Good Judge of Teacher Talent | I Am Glad I Am Participating in the TIF Program | My Job Satisfaction Has Increased due to the TIF Program | I Feel Increased Pressure to Perform due to the TIF Program | I Have Less Freedom to Teach the Way I Would Like to Teach due to the TIF Program | The TIF Program Has Harmed the Collaborative Nature of Teaching | The TIF Program Has Caused Teachers to Work More Effectively | The TIF Program Is Fair | The Process Used to Determine How Bonuses Are Determined Was Adequately Explained to Me | **Number of Teachers** |
| All Teachers (primary analysis) | -4 | -7* | 0.0 | -5 | 0.0 | 14* | 10* | 8* | -6 | -5 | 4 | **784-882** |
| Teaching Assignment | | | | | | | | | | | | |
| (1) Tested grades and subjects | 4 | -6 | 0 | -7 | -2 | 17* | 10 | 8 | -7 | 0 | 6 | **430-481** |
| (2) Nontested grades and subjects | -16* | -10* | -1 | -3 | 2 | 10 | 10 | 8 | -4 | -13 | 1 | **354-401** |
| **Difference between subgroup (1) - (2)** | **20*** | **4** | **1** | **-4** | **-3** | **7** | **0.0** | **0.0** | **-3** | **13** | **6** | |
| Teacher Experience | | | | | | | | | | | | |
| (1) Less than 5 years | -4 | -1 | 12 | -2 | -2 | 11 | 8 | 8 | -6 | -9 | 10 | **157-191** |
| (2) 5 to 15 years | -7 | -10 | -1 | 0 | -1 | 14* | 8 | 5 | -7 | -6 | 2 | **404-448** |
| (3) Greater than 15 years | -2 | -7 | -5 | -15* | 3 | 15* | 11 | 13* | -1 | -1 | 0 | **222-243** |
| **Difference between subgroups (1) - (2)** | **2** | **10** | **13** | **-3** | **-1** | **-3** | **1** | **3** | **1** | **-3** | **8** | |
| **Difference between subgroups (3) - (2)** | **5** | **4** | **-3** | **-15*** | **4** | **1** | **4** | **8** | **5** | **5** | **-2** | |

Source: Teacher survey, 2013.

Note: The difference between the treatment group and the control group is adjusted for block fixed effects and school-level characteristics at time of randomization. All teacher estimates come from Table V.1. Subgroup-specific impact estimates and hypothesis tests are based on a model with a treatment dummy and interaction(s) between the treatment and the subgroup(s) along with main effect for subgroup(s) using the pooled sample.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

## Impacts on Principals' Hiring Autonomy, Staffing, and Compensation Decisions

In this section, we report findings on principals' hiring autonomy, staffing, and compensation decisions. Principals' autonomy in hiring is a necessary, though not sufficient, condition for pay-for-performance to have an effect on principal recruitment strategies. Most principals in both treatment and control schools reported having input in hiring decisions (Table E.8). In addition, the introduction of pay-for-performance in treatment schools may generate incentives for principals to strategically assign teachers to classrooms or use nonmonetary compensation. Because pay-for-performance bonuses depend on students' achievement growth on standardized tests, principals in schools eligible for such bonuses may use different criteria to assign teachers to tested grades and subjects. For example, if school staff can earn a pay-for-performance bonus based on student achievement growth measured at the school level, a principal may decide to assign teachers to tested grades and subjects based on belief in a teacher's ability to raise student achievement scores. Control schools could also compensate for the lack of pay-for-performance bonuses in their schools by making more extensive use of nonmonetary benefits to reward performance, such as giving effective teachers more time for leadership activities or priority in teaching assignments.

**Table E.8. Principals' Autonomy in Hiring Teachers, Cohort 1 (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Impact | Treatment | Control | Impact |
| Principal has complete autonomy over teacher hiring | 12 | 5 | 7* | 22 | 15 | 7 |
| Principal is part of a school-level team responsible for teacher hiring | 52 | 52 | 0 | 47 | 57 | -11 |
| Principal receives a set of prescreened candidates from the district office as the pool from which he or she can interview and hire | 35 | 41 | -5 | 27 | 25+ | 3 |
| Principal has little or no input in hiring teachers at this school | 2 | 3 | -2 | 3 | 2 | 2 |
| **Number of Principals** | **65** | **64** | | **64** | **61** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

**We found no evidence that principals determine teacher assignments or compensations differently in response to pay-for-performance.** Pay-for-performance had no significant impact on most measures of principals' staffing decisions (Table E.9). Treatment and control principals were equally likely to report that they use teacher's ability to produce high test scores when making decisions, suggesting that pay-for-performance is not inducing principals to make strategic assignments of teachers. In Year 2, the only significant different between treatment and control principals is in the use of teacher's seniority when making decisions on teaching assignments.

There is no evidence that control teachers were receiving nonmonetary benefits for not being eligible for pay-for-performance. About 40 percent of principals offer nonmonetary benefits such as release from classroom teaching, increased decision-making authority, or priority in teaching or student assignments. However, there is no difference in the use of nonmonetary benefits between treatment and control principals (Table E.10).

**Table E.9. Criteria Used for Teacher Assignments to Grade Levels or Subject Areas, Cohort 1 (Percentages Who Report They Are "Always" or "Often" Used)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Impact | Treatment | Control | Impact |
| The teacher's experience in a grade level or subject area | 85 | 89 | -4 | 89 | 90 | -1 |
| The teacher's seniority | 3 | 14 | -11* | 13 | 3+ | 9* |
| The teacher's content knowledge | 91 | 97 | -6 | 92 | 93 | -1 |
| The teacher's ability to produce high test scores in grades/classes in which state or federal assessments are administered | 72 | 74 | -2 | 64 | 66 | -2 |
| The teacher's ability to work with certain student populations | 85 | 80 | 6 | 84 | 81 | 3 |
| To balance teacher experience and expertise in a grade level or subject | 72 | 71 | 1 | 69 | 73 | -4 |
| **Number of Principals—Range[a]** | **64-65** | **62-64** | | **62-63** | **58-59** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table E.10. Nonmonetary Benefits Used to Recognize Teachers' Performance or Responsibilities, Cohort 1 (Percentages)**

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Impact | Treatment | Control | Impact |
| Use of nonmonetary benefits | 39 | 38 | 1 | 40 | 37 | 4 |
| Type of nonmonetary benefits: | | | | | | |
| Release from classroom teaching for mentoring or other leadership activities | 35 | 27 | 9 | 28 | 32 | -3 |
| Decision-making authority on issues such as hiring staff or adopting curriculum | 31 | 28 | 3 | 32 | 30 | 2 |
| Priority in teaching assignments | 7 | 11 | -4 | 9 | 18 | -9 |
| Priority in student assignments | 4 | 3 | 1 | 3 | 7 | -4 |
| **Number of Principals—Range[a]** | **64** | **64** | | **63-64** | **60** | |

Source: Principal survey, 2012 and 2013.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

+Difference between Year 1 and Year 2 within treatment status is statistically significant at the .05 level, two-tailed test.

THIS PAGE IS INTENTIONALLY BLANK

# APPENDIX F

# SUPPLEMENTAL FINDINGS ON IMPACTS OF PAY-FOR-PERFORMANCE ON EDUCATOR EFFECTIVENESS AND STUDENT ACHIEVEMENT FOR CHAPTER VI

THIS PAGE IS INTENTIONALLY BLANK

This appendix supplements the findings presented in Chapter VI that examined impacts of pay-for-performance on educator effectiveness and student achievement.

As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts whose schools were randomly assigned in spring and summer 2011 were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 1 completed two years of implementation during the period of this study, 2011–2012 and 2012–2013, referred to as Years 1 and 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort.

This appendix includes findings for Cohorts 1 and 2, supplemental findings for Cohort 1 (for example, subgroup findings), and sensitivity analyses that assess the robustness of the main impact estimates reported in Chapter VI.

## Educator Performance Ratings

This section presents two types of additional analyses of the impact of pay-for-performance on educator performance ratings: (1) sensitivity analyses that assess the robustness of the main impact estimates and (2) findings that include both Cohorts 1 and 2.

### Sensitivity Analyses

Tables F.1 and F.2 explore the sensitivity of the main impact estimates for school achievement growth ratings and teacher observation ratings to several changes to the regression model or estimation sample, described below.

**Table F.1. Impacts of Pay-for-Performance on School Achievement Growth Ratings Using Alternative Specifications, Cohort 1**

| Time Period and Model | Treatment Schools | Control Schools | Impact | *P*-value | **Number of Schools** |
|---|---|---|---|---|---|
| **Year 1** | | | | | |
| **Main Model** | 2.59 | 2.25 | 0.34* | 0.046 | **124** |
| **Alternative Specifications** | | | | | |
| Weights | | | | | |
| (1) Districts are weighted equally | 2.56 | 2.23 | 0.34* | 0.020 | **124** |
| Covariates | | | | | |
| (2) No covariates except randomization block indicators | 2.52 | 2.25 | 0.27 | 0.098 | **124** |
| **Year 2** | | | | | |
| **Main Model** | 2.46 | 2.21 | 0.25* | 0.047 | **131** |
| **Alternative Specifications** | | | | | |
| Weights | | | | | |
| (1) Districts are weighted equally | 2.51 | 2.27 | 0.24 | 0.114 | **131** |
| Covariates | | | | | |
| (2) No covariates except randomization block indicators | 2.40 | 2.21 | 0.19 | 0.144 | **131** |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

*Impact is statistically significant at the .05 level, two-tailed test.

**Table F.2. Impacts of Pay-for-Performance on Teachers' Classroom Observation Ratings Using Alternative Specifications, Cohort 1**

| Time Period and Model | Teachers in Treatment Schools | Teachers in Control Schools | Impact | *p*-value | **Number of Teachers** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Year 1** | | | | | | |
| **Main Model** | 2.94 | 2.91 | 0.03 | 0.243 | **3,625** | **132** |
| **Alternative Specifications** | | | | | | |
| Weights | | | | | | |
| (1) Teachers are weighted equally | 2.92 | 2.88 | 0.04 | 0.051 | **3,625** | **132** |
| (2) Districts are weighted equally | 2.89 | 2.84 | 0.05 | 0.095 | **3,625** | **132** |
| Covariates | | | | | | |
| (3) No covariates except randomization block indicators | 2.95 | 2.91 | 0.03 | 0.110 | **3,625** | **132** |
| Unit of Analysis | | | | | | |
| (4) All data are averaged to the cluster level | 2.96 | 2.91 | 0.05 | 0.117 | **NA** | **90[a]** |
| **Year 2** | | | | | | |
| **Main Model** | 3.02 | 2.97 | 0.05 | 0.070 | **3,628** | **132** |
| **Alternative Specifications** | | | | | | |
| Weights | | | | | | |
| (1) Teachers are weighted equally | 2.98 | 2.92 | 0.06* | 0.010 | **3,628** | **132** |
| (2) Districts are weighted equally | 2.97 | 2.91 | 0.06* | 0.031 | **3,628** | **132** |
| Covariates | | | | | | |
| (3) No covariates except randomization block indicators | 3.01 | 2.97 | 0.05 | 0.093 | **3,628** | **132** |
| Unit of Analysis | | | | | | |
| (4) All data are averaged to the cluster level | 3.04 | 2.97 | 0.08 | 0.075 | **NA** | **90[a]** |

Source: Educator administrative data.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

[a]Sample size denotes the number of clusters. Some clusters had multiple schools.

*Impact is statistically significant at the .05 level, two-tailed test.

NA is not applicable.

**Using alternate weighting approaches.** In our main specification, we normalized the analysis weights so that each school received the same weight in the final analysis sample. Therefore, in the main impact estimates, districts with more schools received more weight than those with fewer schools. In addition, for the teacher observation ratings, teachers in large schools received less weight than those in small schools. We explored two alternative approaches to normalizing sample weights. In the first alternative approach (for analyses of school achievement growth ratings and teacher observation ratings), each district received the same weight. This approach produced estimates of the impact of pay-for-performance in the average Cohort 1 district, which could be of interest because each district designed its TIF program in a different way. In the second alternative approach (for analyses of teacher observation ratings), each teacher received the same weight. This approach produced estimates of the impact of pay-for-performance on the average teacher, which could be of interest because pay-for-performance was intended to change teachers' behavior.

For school achievement growth ratings, estimates from the model that gave districts equal weight were similar to those from the main model, except that the impact for Year 2 was not significant (Table F.1, model 1). In contrast, although there was no impact of pay-for-performance on teacher observation ratings when schools were weighted equally, of the four estimates using these alternative weighting approaches, all were positive, two were significant, and two were almost significant (*p*-values less than 0.10; Table F.2, models 1 and 2).

**Excluding covariates.** Our main estimation model controlled for randomization block indicators and the school-level pre-implementation means of student achievement and student race/ethnicity. Controlling for schools' pre-implementation characteristics accounted for the fact that treatment schools had slightly lower student math achievement and slightly different student racial/ethnic composition than control schools at the beginning of the study. Failure to account for these preexisting differences could generate an inaccurate estimate of the effects of pay-for-performance. Nevertheless, because some researchers have expressed methodological concerns about the use of covariates in analyzing experimental data (Freedman 2008), we also estimated a model that included no other covariates besides the randomization block indicators. As expected, excluding covariates reduced the precision of the estimates, resulting in *p*-values slightly greater than the main model. For school achievement growth ratings, in contrast to the main model, this specification found no significant impact of pay-for-performance in either year (Table F.1, model 2). For teacher observation ratings, neither the main model nor this specification found significant impacts of pay-for-performance (Table F.2, model 3).

**Using clusters as the unit of analysis.** The main specification for teacher observation ratings used teachers as the unit of analysis and used robust standard errors that accounted for the clustering of teachers' outcomes within the clusters (schools or groups of schools) that were assigned to the treatment and control groups. Because clustered standard errors can be biased with finite numbers of clusters (Donald and Lang 2007), we explored an alternative model that used cluster-level averages of the dependent and independent variables to avoid the use of cluster-robust standard errors. Findings from this model were similar to the main findings (Table F.2, model 4).

**Findings for Cohorts 1 and 2**

In Tables F.3 and F.4, we present the impact of pay-for-performance on the Year 1 performance ratings of educators in schools in Cohorts 1 and 2, as well as the main impact estimates from Chapter VI, which only included educators in Cohort 1 schools. Unlike estimates based on only Cohort 1, the estimated impacts of pay-for-performance on school achievement growth ratings and classroom achievement growth ratings in Year 1 were no longer found to be significant (*p*-values = 0.06) when

both cohorts were included in the analysis. Estimated impacts on observation ratings in Cohorts 1 and 2 were similar to those in Cohort 1 only.

**Table F.3. Student Achievement Growth Ratings in Year 1, Cohorts 1 and 2**

| | Treatment | Control | Impact | *p*-value | **Number of Teachers** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Cohort 1** | | | | | | |
| Ratings for Student Achievement Growth in Schools | 2.59 | 2.25 | 0.34* | 0.046 | NA | **124** |
| Ratings for Student Achievement Growth in Classrooms | 2.26 | 2.08 | 0.18* | 0.033 | **1,093** | **73** |
| **Cohorts 1 and 2** | | | | | | |
| Ratings for Student Achievement Growth in Schools | 2.22 | 1.98 | 0.24 | 0.062 | NA | **174** |
| Ratings for Student Achievement Growth in Classrooms | 2.02 | 1.91 | 0.11 | 0.060 | **2,439** | **118** |

Source: Educator administrative data.

Notes: School achievement growth ratings for one district in Year 1 are omitted because they could not be converted to a 1 to 4 rating scale. This district awarded school-level bonuses in Year 1 based on schools' relative rank among schools in the district on school achievement growth (for example, which schools had the most growth), so there were no theoretical minimums or maximums for these measures. Classroom achievement growth ratings are only available for the six districts in Cohort 1 and three districts in Cohort 2 that evaluated teachers based on classroom achievement growth. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

NA is not applicable

*Impact is statistically significant at the .05 level, two-tailed test.

**Table F.4. Observation Ratings for Teachers and Principals in Year 1, Cohorts 1 and 2**

| | Treatment | Control | Impact | *p*-value | **Number of Educators** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Cohort 1** | | | | | | |
| Teachers' Classroom Observation Ratings | 2.94 | 2.91 | 0.03 | 0.243 | **3,625** | **132** |
| Observation Ratings for Principals | 3.08 | 3.18 | -0.10 | 0.197 | **105** | **105** |
| **Cohorts 1 and 2** | | | | | | |
| Teachers' Classroom Observation Ratings | 3.15 | 3.12 | 0.03 | 0.075 | **5,219** | **183** |
| Observation Ratings for Principals | 3.26 | 3.40 | -0.14 | 0.053 | **151** | **151** |

Source: Educator administrative data.

Notes: None of the impacts were statistically significant at the .05 level. One district did not provide observation ratings for principals in Year 1. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

## Retention and Recruitment of Effective Educators

In Chapter VI, we examined whether pay-for-performance led to the retention and recruitment of more higher-performing educators. This section presents supplemental analyses of whether pay-for-performance led to staffing changes at schools that offered performance bonuses. First, we contextualize the main findings by examining overall retention rates among teachers and principals at treatment and control schools. Second, we present evidence on whether pay-for-performance led to a change in the average professional and demographic characteristics of educators working at the schools. Third, we present additional findings on the retention and recruitment of effective educators based on alternative definitions of staying, leaving, and entering the study schools.

### Overall Retention Rates

Overall retention rates—that is, percentages of educators who stayed in their schools between years—provide important context for analyzing whether pay-for-performance triggered staffing changes that resulted in more higher-performing educators working at these schools. As discussed in Chapter VI, the extent of educator turnover at a school determines how much scope there is for staffing changes to shape the overall effectiveness of the school's staff. For example, if a large school only had one teacher depart each year, then overall effectiveness would change little if the departing teacher was the worst teacher rather than the best. Likewise, the effectiveness of the departing teacher's replacement would have little influence on overall effectiveness at the school.

We measured retention for all full-time teachers and principals working in study schools in Year 1. Educators were considered retained if they returned to the same school and position (teacher or principal) in the fall of Year 2 (one-year retention) and the fall of Year 3 (two-year retention). We also measured one-year retention for all full-time educators working in study schools in Year 2. Differences in retention rates between treatment and control schools measured the impact of pay-for-performance on educator retention.

In the study schools, about one-fifth of teachers departed between consecutive years, and one-third of teachers departed over a two-year period (Table F.5). Likewise, about one-fifth to one-fourth of principals departed between consecutive years, and two-fifths of principals departed over a two-year period (Table F.6). Therefore, although many educators were retained, there was also plenty of turnover, leaving the potential for staffing changes to be an important way of shaping educator effectiveness.

We found no impact of pay-for-performance on the overall retention rates of either teachers or principals. This implies that any increases in the retention of higher-performing educators as a result of pay-for-performance should have been offset by decreases in the retention of lower-performing educators. In fact, evidence from Chapter VI suggested that pay-for-performance caused more higher-performing principals to stay in their schools and more lower-performing principals to leave their schools.

**Table F.5. Teachers Who Continued Teaching in the Same School Across Multiple Years, Cohort 1 (Percentages)**

| Period | Treatment | Control | Impact | P-value | **Number of Teachers** | **Number of Schools** |
|---|---|---|---|---|---|---|
| One-Year Period | | | | | | |
| Between Years 1 and 2 | 81 | 80 | 1 | 0.447 | **4,346** | **132** |
| Between Years 2 and 3 | 78 | 77 | 1 | 0.501 | **4,466** | **132** |
| Two-Year Period | | | | | | |
| Between Years 1 and 3 | 65 | 64 | 2 | 0.223 | **4,346** | **132** |

Source: Educator administrative data.

Notes: None of the impacts were statistically significant at the .05 level. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

**Table F.6. Principals Who Continued Leading the Same School Across Multiple Years, Cohort 1 (Percentages)**

| Period | Treatment | Control | Impact | P-value | **Number of Principals** | **Number of Schools** |
|---|---|---|---|---|---|---|
| One-Year Period | | | | | | |
| Between Years 1 and 2 | 80 | 73 | 7 | 0.273 | **134** | **128** |
| Between Years 2 and 3 | 79 | 80 | -2 | 0.833 | **138** | **129** |
| Two-Year Period | | | | | | |
| Between Years 1 and 3 | 67 | 58 | 9 | 0.363 | **134** | **128** |

Source: Educator administrative data.

Notes: None of the impacts were statistically significant at the .05 level. The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

## Impacts of Pay-for-Performance on Other Characteristics of Schools' Staff

Given that pay-for-performance was intended to help schools retain and attract more effective educators, any staffing changes resulting from pay-for-performance could have also altered other characteristics of the schools' staff, including the demographic and professional characteristics of teachers and principals. However, we found no evidence that pay-for-performance led to changes in those staff characteristics. In Year 2, educators working in treatment and control schools had similar demographic characteristics and professional background (Table F.7).

**Table F.7. Characteristics of Teachers and Principals in Year 2, Cohort 1 (Percentages Unless Otherwise Noted)**

| | Teachers | | | Principals | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| Demographic Characteristics | | | | | | |
| Female | 85 | 84 | 1 | 62 | 68 | -5 |
| Race/Ethnicity | | | | | | |
| White, non-Hispanic | 73 | 72 | 1 | 60 | 52 | 8 |
| Black, non-Hispanic | 19 | 22 | -2 | 31 | 36 | -5 |
| Hispanic | 2 | 2 | 0 | 3 | 3 | 0 |
| Other | 5 | 4 | 1 | 5 | 8 | -3 |
| Age (average years) | 42 | 41 | 0 | 48 | 49 | -1 |
| Education | | | | | | |
| Master's degree or higher | 50 | 51 | -2 | 97 | 97 | 0 |
| Experience in K–12 Education | | | | | | |
| Total experience (average years) | 11 | 11 | 0 | 15 | 14 | 1 |
| Less than 5 years | 27 | 28 | -1 | 19 | 20 | 0 |
| 5-15 years | 45 | 45 | 0 | 34 | 43 | -8 |
| More than 15 years | 28 | 27 | 1 | 46 | 38 | 8 |
| **Number of Educators—Range[a]** | **1,533-2,140** | **1,618-2,208** | | **51-63** | **47-65** | |
| **Number of Schools—Range[a]** | **50-66** | **50-66** | | **48-60** | **44-61** | |

Source: Educator administrative data.

Notes: None of the differences between treatment and control educators were statistically significant at the .05 level. The difference between the treatment and control estimates may not equal the difference shown in the table due to rounding.

## Impacts of Pay-for-Performance on the Retention and Recruitment of Effective Educators Using Alternative Definitions of Staying, Leaving, and Entering Study Schools

In Chapter VI, we examined differences in performance ratings between educators who stayed in treatment schools and those who stayed in control schools, and between educators who left treatment schools and those who left control schools (see Figures VI.1 and VI.3). In those main analyses, educators who worked in study schools in Year 1 were classified as having subsequently stayed in or left their schools based on whether they continued working in the same schools and positions between Year 1 and the fall of Year 3. We measured the effectiveness of each group with Year 1 performance ratings.

Tables F.8 and F.9 show the effectiveness of educators who stayed in and left their schools based on alternative time periods for assessing performance and measuring retention. In particular, we examined the Year 1 performance ratings of educators who stayed in and left their schools between Years 1 and 2, and the Year 2 performance ratings of educators who stayed in and left their schools between Years 2 and 3. For teachers, findings from these alternative time periods were similar to the main findings: we found no evidence that pay-for-performance led to more higher-performing teachers deciding to stay in their schools or more lower-performing teachers deciding to leave their schools. For principals, both the main findings and those from alternative time periods indicate that principals who stayed in treatment schools earned higher school achievement growth ratings than principals who stayed in control schools. However, although the main findings from Chapter VI indicate that principals who left treatment schools had lower observation ratings than principals who left control schools, there was no statistically significant difference between these groups in the alternative time periods.

**Table F.8. Classroom Observation and Classroom Achievement Growth Ratings of Teachers Who Stayed in and Left Their Schools Between Consecutive Years, Cohort 1 (Points on 1 to 4 Scale)**

| | Teachers Who Stayed Between Years 1 and 2 | | Teachers Who Left Between Years 1 and 2 | |
|---|---|---|---|---|
| Outcomes Measured in Year 1 | Treatment | Control | Treatment | Control |
| Classroom Observation Rating | 2.96 | 2.95 | 2.88 | 2.81 |
| Classroom Achievement Growth Rating | 2.30 | 2.11 | 2.18 | 2.05 |
| **Number of Teachers** | **1,719** | **1,711** | **470** | **446** |
| **With classroom observation rating** | **1,514** | **1,476** | **314** | **321** |
| **With classroom achievement growth rating** | **445** | **408** | **110** | **130** |
| | Teachers Who Stayed Between Years 2 and 3 | | Teachers Who Left Between Years 2 and 3 | |
| Outcomes Measured in Year 2 | Treatment | Control | Treatment | Control |
| Classroom Observation rating | 3.05 | 3.00 | 2.91 | 2.89 |
| Classroom Achievement Growth Rating | 2.28 | 2.23 | 2.18 | 2.22 |
| **Number of Teachers** | **1,702** | **1,713** | **505** | **546** |
| **With classroom observation rating** | **1,488** | **1,471** | **312** | **357** |
| **With classroom achievement growth rating** | **573** | **576** | **85** | **108** |

Source: Educator administrative data.

Note: None of the differences between teachers in treatment and control schools were statistically significant at the .05 level.

**Table F.9. Observation and School Achievement Growth Ratings of Principals who Stayed in and Left Their Schools Between Consecutive Years, Cohort 1 (Points on 1 to 4 Scale)**

| | Principals Who Stayed Between Years 1 and 2 | | Principals Who Left Between Years 1 and 2 | |
|---|---|---|---|---|
| Outcomes Measured in Year 1 | Treatment | Control | Treatment | Control |
| Observation Rating | 3.09 | 3.16 | 2.90 | 3.18 |
| School Achievement Growth Rating | 2.68* | 2.19 | 2.19 | 2.42 |
| **Number of Principals** | **50** | **50** | **15** | **19** |
| **With observation rating** | **43** | **40** | **10** | **12** |
| **With school achievement growth rating** | **48** | **48** | **14** | **17** |
| | Principals Who Stayed Between Years 2 and 3 | | Principals Who Left Between Years 2 and 3 | |
| Outcomes Measured in Year 2 | Treatment | Control | Treatment | Control |
| Observation Rating | 3.24 | 3.10 | 2.98 | 2.83 |
| School Achievement Growth Rating | 2.59* | 2.15 | 2.13 | 2.64 |
| **Number of Principals** | **52** | **54** | **16** | **16** |
| **With observation rating** | **49** | **46** | **12** | **11** |
| **With school achievement growth rating** | **52** | **53** | **16** | **16** |

Source: Educator administrative data.

*Difference between principals of treatment and control schools is statistically significant at the .05 level, two-tailed test.

In Chapter VI, we also examined whether pay-for-performance caused more higher-performing educators to be hired at schools that offered performance bonuses. To answer this question, our main analyses compared the Year 2 performance ratings of treatment and control educators who were new to their schools in that year. As discussed in Chapter VI, we focused on new recruits in Year 2 because educators' decisions on where to work in Year 2 could have been shaped by districts' and schools' efforts in Year 1 to make educators aware of the TIF program. However, because schools were randomly assigned to the treatment and control group in the spring and summer before Year 1, it is possible that pay-for-performance could have enabled schools to recruit better educators to begin working in Year 1.

For teachers, we found no evidence that new recruits in Year 1 were more effective in treatment schools than control schools (Table F.10). When examining newly hired principals in Year 1, we found that those in treatment schools earned higher school achievement growth ratings in that year than those in control schools (Table F.11). Although this finding may suggest that pay-for-performance led to the recruitment of more effective principals in Year 1, it is unclear whether schools' eligibility for pay-for-performance would have been known to prospective principals at the time of hire. An alternative explanation for this finding is that pay-for-performance could have motivated newly hired principals in treatment schools to work more effectively than their counterparts in control schools in Year 1. A third explanation is that positive impacts on school achievement growth ratings could have also reflected improvements by teachers and other school staff—not just principals.

**Table F.10. Classroom Observation and Classroom Achievement Growth Ratings of Teachers Who Were New to Their Schools in Year 1, Cohort 1 (Points on 1 to 4 Scale)**

| Outcomes Measured in Year 1 | Treatment | Control |
|---|---|---|
| Classroom Observation Rating | 2.89 | 2.87 |
| Classroom Achievement Growth Rating | 2.14 | 2.20 |
| **Number of Teachers** | **355** | **389** |
| **With classroom observation rating** | **292** | **323** |
| **With classroom achievement growth rating** | **81** | **80** |

Source: Educator administrative data.

Note: None of the differences between teachers in treatment and control schools were statistically significant at the .05 level.

**Table F.11. Observation and School Achievement Growth Ratings of Principals Who Were New to Their Schools in Year 1, Cohort 1 (Points on 1 to 4 Scale)**

| Outcomes Measured in Year 1 | Treatment | Control |
|---|---|---|
| Observation Rating | 3.15 | 2.98 |
| School Achievement Growth Rating | 2.77* | 1.91 |
| **Number of Principals** | **10** | **11** |
| **With observation rating** | **10** | **11** |
| **With school achievement growth rating** | **10** | **11** |

Source: Educator administrative data.

*Difference between principals of treatment and control schools is statistically significant at the .05 level, two-tailed test.

## Student Achievement

This section presents three types of additional analyses of the impacts of pay-for-performance on student achievement: (1) sensitivity analyses that assess the robustness of the main impact estimates, (2) findings that include both Cohorts 1 and 2, and (3) subgroup analyses that assess impacts within elementary and middle grades separately.

### Sensitivity Analyses

We explored the sensitivity of the main impact estimates to several changes to the regression model or estimation sample (Tables F.12 through F.15). Findings from these specifications were generally similar to the main impact estimates, with some exceptions described below.

**Table F.12. Impacts of Pay-for-Performance on Student Achievement in Reading Using Alternate Specifications in Year 1, Cohort 1**

| | Impact (student *z*-score units) | *P*-value | **Number of Students** | **Number of Schools** |
|---|---|---|---|---|
| **Main Model** | 0.03* | 0.040 | **40,576** | **132** |
| **Alternative Specifications** | | | | |
| Standardizing Test Scores | | | | |
| (1) Compute z-scores using sample means/standard deviations | 0.04* | 0.039 | **40,576** | **132** |
| Weights | | | | |
| (2) Students weighted equally | 0.03* | 0.040 | **40,576** | **132** |
| (3) Districts weighted equally | 0.03* | 0.026 | **40,576** | **132** |
| Sample of Students | | | | |
| (4) Only include grades with pretests[a] | 0.05* | 0.004 | **33,644** | **130[b]** |
| Covariates | | | | |
| (5) No covariates except randomization block indicators | 0.00 | 0.893 | **40,576** | **132** |
| (6) Only covariates are school-level pre-implementation means of student achievement and student race/ethnicity and randomization block indicators | 0.03 | 0.059 | **40,576** | **132** |
| (7) All covariates interacted with state indicators | 0.05* | 0.000 | **40,576** | **132** |
| (8) Include student pretests interacted with grade indicators | 0.03* | 0.043 | **40,576** | **132** |
| (9) Include student pretests, squared and cubed | 0.03* | 0.043 | **40,576** | **132** |
| Unit of Analysis | | | | |
| (10) All data are averaged to the cluster level and the only covariates are those in model (6) | 0.04 | 0.120 | **NA** | **90[c]** |

Source: Student administrative data.

[a]Grades with pretests are grades 4 through 8 in Year 1 and grades 5 through 8 in Year 2.

[b]The excluded schools serve students in grades K–3.

[c]Sample size denotes the number of clusters. Some clusters had multiple schools.

*Impact is statistically significant at the .05 level, two-tailed test.

NA is not applicable

**Table F.13. Impacts of Pay-for-Performance on Student Achievement in Reading Using Alternate Specifications in Year 2, Cohort 1**

| | Impact (student z-score units) | *P*-value | **Number of Students** | **Number of Schools** |
|---|---|---|---|---|
| **Main Model** | 0.03* | 0.026 | **40,391** | **132** |
| **Alternative Specifications** | | | | |
| Standardizing Test Scores | | | | |
| (1) Compute z-scores using sample means/standard deviations | 0.03* | 0.026 | **40,391** | **132** |
| Weights | | | | |
| (2) Students weighted equally | 0.03* | 0.008 | **40,391** | **132** |
| (3) Districts weighted equally | 0.02 | 0.206 | **40,391** | **132** |
| Sample of Students | | | | |
| (4) Only include grades with pretests[a] | 0.05* | 0.004 | **27,136** | **129[b]** |
| Covariates | | | | |
| (5) No covariates except randomization block indicators | 0.00 | 0.874 | **40,391** | **132** |
| (6) Only covariates are school-level pre-implementation means of student achievement and student race/ethnicity and randomization block indicators | 0.03* | 0.013 | **40,391** | **132** |
| (7) All covariates interacted with state indicators | 0.04* | 0.012 | **40,391** | **132** |
| (8) Include student pretests interacted with grade indicators | 0.03* | 0.024 | **40,391** | **132** |
| (9) Include student pretests, squared and cubed | 0.03* | 0.028 | **40,391** | **132** |
| Unit of Analysis | | | | |
| (10) All data are averaged to the cluster level and the only covariates are those in model (6) | 0.03 | 0.148 | **NA** | **90[c]** |

Source: Student administrative data.

[a]Grades with pretests are grades 4 through 8 in Year 1 and grades 5 through 8 in Year 2.

[b]The excluded schools serve students in grades K–3 or K–4.

[c]Sample size denotes the number of clusters. Some clusters had multiple schools.

*Impact is statistically significant at the .05 level, two-tailed test.

NA is not applicable.

**Table F.14. Impacts of Pay-for-Performance on Student Achievement in Math Using Alternate Specifications in Year 1, Cohort 1**

| | Impact (student z-score units) | *P*-value | **Number of Students** | **Number of Schools** |
|---|---|---|---|---|
| **Main Model** | 0.02 | 0.335 | **40,852** | **132** |
| **Alternative Specifications** | | | | |
| Standardizing Test Scores | | | | |
| (1) Compute z-scores using sample means/standard deviations | 0.02 | 0.296 | **40,852** | **132** |
| Weights | | | | |
| (2) Students weighted equally | 0.02 | 0.371 | **40,852** | **132** |
| (3) Districts weighted equally | 0.02 | 0.225 | **40,852** | **132** |
| Sample of Students | | | | |
| (4) Only include grades with pretests[a] | 0.03 | 0.168 | **33,819** | **130[b]** |
| Covariates | | | | |
| (5) No covariates except randomization block indicators | -0.03 | 0.319 | **40,852** | **132** |
| (6) Only covariates are school-level pre-implementation means of student achievement and student race/ethnicity and randomization block indicators | 0.02 | 0.352 | **40,852** | **132** |
| (7) All covariates interacted with state indicators | 0.04 | 0.100 | **40,852** | **132** |
| (8) Include student pretests interacted with grade indicators | 0.02 | 0.327 | **40,852** | **132** |
| (9) Include student pretests, squared and cubed | 0.02 | 0.345 | **40,852** | **132** |
| Unit of Analysis | | | | |
| (10) All data are averaged to the cluster level and the only covariates are those in model (6) | 0.02 | 0.498 | **NA** | **90[c]** |

Source: Student administrative data.

Note: None of the impacts were statistically significant at the .05 level, two-tailed test.

[a]Grades with pretests are grades 4 through 8 in Year 1 and grades 5 through 8 in Year 2.

[b]The excluded schools serve students in grades K–3.

[c]Sample size denotes the number of clusters. Some clusters had multiple schools.

NA is not applicable.

**Table F.15. Impacts of Pay-for-Performance on Student Achievement in Math Using Alternate Specifications in Year 2, Cohort 1**

| | Impact (student z-score units) | *P*-value | **Number of Students** | **Number of Schools** |
|---|---|---|---|---|
| **Main Model** | 0.04 | 0.068 | **40,709** | **132** |
| **Alternative Specifications** | | | | |
| Standardizing Test Scores | | | | |
| (1) Compute z-scores using sample means/standard deviations | 0.05 | 0.053 | **40,709** | **132** |
| Weights | | | | |
| (2) Students weighted equally | 0.04 | 0.060 | **40,709** | **132** |
| (3) Districts weighted equally | 0.04 | 0.150 | **40,709** | **132** |
| Sample of Students | | | | |
| (4) Only include grades with pretests[a] | 0.06* | 0.025 | **27,292** | **129[b]** |
| Covariates | | | | |
| (5) No covariates except randomization block indicators | 0.01 | 0.598 | **40,709** | **132** |
| (6) Only covariates are school-level pre-implementation means of student achievement and student race/ethnicity and randomization block indicators | 0.05* | 0.032 | **40,709** | **132** |
| (7) All covariates interacted with state indicators | 0.05* | 0.044 | **40,709** | **132** |
| (8) Include student pretests interacted with grade indicators | 0.04 | 0.068 | **40,709** | **132** |
| (9) Include student pretests, squared and cubed | 0.04 | 0.071 | **40,709** | **132** |
| Unit of Analysis | | | | |
| (10) All data are averaged to the cluster level and the only covariates are those in model (6) | 0.05 | 0.111 | **NA** | **90[c]** |

Source: Student administrative data.

[a]Grades with pretests are grades 4 through 8 in Year 1 and grades 5 through 8 in Year 2.

[b]The excluded schools serve students in grades K–3 or K–4.

[c]Sample size denotes the number of clusters. Some clusters had multiple schools.

*Impact is statistically significant at the .05 level, two-tailed test.

NA is not applicable.

**Standardizing test scores.** For the main analysis, we standardized outcome and baseline test scores into $z$-scores based on grade-specific means and standard deviations of test scores in each statewide population. We explored an alternative method of standardizing test scores into $z$-scores based on the grade-specific means and standard deviations of test scores for students in control schools in the same state. Findings from these specifications were similar to the main impact estimates (Tables F.12 through F.15, model 1).

**Using alternate weighting approaches.** In our main specification, we normalized the analysis weights that each school received the same weight in the final analysis sample. Therefore, in the main impact estimates, students in large schools received less weight than those in small schools, and districts with more schools received more weight than those with fewer schools. We explored two alternative approaches to normalizing sample weights. In the first alternative approach, each district received the same weight. This approach produced estimates of the impact of pay-for-performance in the average Cohort 1 district, which could be of interest because each district designed its TIF program

in a different way. In the second alternative approach, each student received the same weight. This approach produced estimates of the impact of pay-for-performance on the average student, which could be of interest because pay-for-performance was ultimately intended to improve student outcomes. Findings from these models were similar to the main impact estimates (Tables F.12 through F.15, models 2 and 3), with one exception. The positive impact of pay-for-performance on reading in Year 2 was not significant when we gave each district the same weight (Table F.13, model 3).

**Using an alternate sample of students.** Our main analysis included students in grades 3 through 8 and controlled for pretest scores from the pre-implementation year. Because the assessments were administered in grades 3 through 8, 3rd graders in Year 1 and 3rd and 4th graders in Year 2 were missing pretest scores from the pre-implementation year. For the main analyses presented in Chapter VI, these students were assigned placeholder values for their pretest scores, and the regression models controlled for indicators of missing pretest scores (see Appendix B). When we excluded these grades from the analyses, the findings were similar to the main impact estimates (Tables F.12 through F.15, model 4). The one exception is that when we excluded grades 3 and 4, the estimated impact of pay-for-performance on math achievement in Year 2 was positive and statistically significant (Table F.15, model 4).

**Changing covariates.** Our main estimation model controlled for randomization block indicators and the student- and school-level covariates described in Appendix B. To assess the sensitivity of the estimates to the choice of covariates or the method of controlling for pretest scores, we estimated several alternative models.

First, we omitted all covariates except the randomization block indicators (Tables F.12 through F.15, model 5). Unlike this alternative model, the main model controlled for schools' pre-implementation characteristics (along with student-level covariates) to account for the fact that treatment schools had slightly lower student math achievement and slightly different student racial/ethnic composition than control schools at the beginning of the study. Failure to account for these preexisting differences could generate an inaccurate estimate of the effects of pay-for-performance. Nevertheless, because some researchers have expressed methodological concerns about the use of covariates in analyzing experimental data (Freedman 2008), we estimated this alternative model, dropping all covariates besides the randomization block indicators. As expected, when we did not account for preexisting differences between treatment and control schools, the alternative estimates differed from our main findings. For reading in both years, although the main model found a statistically significant impact of 0.03, the impact from the alternative model was a statistically insignificant 0.00. For math, both the main and alternative model found statistically insignificant impacts in both years, but the impacts from the alternative model were smaller than those from the main model in Year 1 (-0.03 versus 0.02) and Year 2 (0.01 versus 0.04).

Second, we omitted student-level covariates—those measuring the individual characteristics of students in the analysis sample—but included randomization block indicators and school-level pre-implementation means of student achievement and student race/ethnicity (Tables F.12 through F.15, model 6). Because pay-for-performance could have affected families' decisions on where to enroll their children and, thus, the characteristics of a school's student population, omitting student-level covariates could avoid biases from controlling for factors that might have been influenced by pay-for-performance. This model produced impact estimates that were similar in magnitude to, but sometimes different in statistical significance from, the main estimates. In particular, the estimated impact on reading achievement in Year 1 was not significant in this model (whereas it was significant in the main model), and the estimated impact on math achievement in Year 2 was significant in this model (whereas it was not significant in the main model).

We also explored models that permitted more flexible functional forms for the covariates. These models differed from the main model in that they (1) added interactions between all covariates in the main estimation model and state indicators, (2) added interactions between the student pretest scores and grade indicators, or (3) included a cubic polynomial of student pretests. The findings from these models were, in general, similar to the main impact estimates (Tables F.12 through F.15, models 7 through 9).

**Using clusters as the unit of analysis.** The main specification used students as the unit of analysis and used robust standard errors that accounted for the clustering of students' outcomes within the clusters (schools or groups of schools) that were assigned to the treatment and control groups. Because clustered standard errors can be biased with finite numbers of clusters (Donald and Lang 2007), we explored an alternative model that used cluster-level averages of the dependent and independent variables to avoid the use of cluster-robust standard errors. Due to the limited number of clusters, this model used a parsimonious set of covariates consisting of the randomization block indicators and the school-level pre-implementation means of student achievement and student race/ethnicity. The estimated impacts from this alternative model were similar in magnitude to those from the main model, but the *p*-values were higher (Table F.12 through F.15, model 10). Therefore, the estimated impacts of pay-for-performance on reading achievement were no longer statistically significant in this alternative model.

### Findings for Cohorts 1 and 2

In Table F.16, we present the impact of pay-for-performance on math and reading achievement in Year 1 for Cohorts 1 and 2, as well as the main impact estimates from Chapter VI, which only included Cohort 1 schools. When Cohort 2 schools were included in the analysis, pay-for-performance no longer had a significant impact on achievement in reading in Year 1. Impacts for math were not significant in Year 1 in either the sample that only included Cohort 1 or the sample that included both cohorts.

**Table F.16. Student Achievement in Math and Reading in Year 1, Cohorts 1 and 2 (Student *z*-score units)**

| Cohort and Subject | Treatment | Control | Impact | *P*-value | **Number of Students** | **Number of Schools** |
|---|---|---|---|---|---|---|
| **Cohort 1** | | | | | | |
| Math | -0.43 | -0.45 | 0.02 | 0.335 | **40,852** | **132** |
| Reading | -0.37 | -0.40 | 0.03* | 0.040 | **40,576** | **132** |
| **Cohorts 1 and 2** | | | | | | |
| Math | -0.55 | -0.56 | 0.01 | 0.327 | **56,566** | **183** |
| Reading | -0.51 | -0.52 | 0.02 | 0.197 | **56,067** | **183** |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

*Impact is statistically significant at the .05 level, two-tailed test.

### Subgroup Findings

In Table F.17, we present the impacts of pay-for-performance on student achievement separately within elementary grades (grades 3 through 5) and middle grades (grades 6 through 8). For elementary and middle school students, impacts on reading scores in Year 2 were nearly identical to the overall impact but were not statistically significant. Across all grade spans, subjects, and years, the only statistically significant impact of pay-for-performance was a positive impact on the reading scores of middle school students in Year 1.

**Table F.17. Student Achievement in Math and Reading in Elementary and Middle Grades, Cohort 1 (Student z-score units)**

| | Math | | | | Reading | | | |
|---|---|---|---|---|---|---|---|---|
| Year and Grades | Treatment | Control | Impact | *p*-value | Treatment | Control | Impact | *p*-value |
| **Year 1** | | | | | | | | |
| Grades 3–5 | -0.44 | -0.45 | 0.00 | 0.859 | -0.40 | -0.42 | 0.02 | 0.270 |
| Grades 6–8 | -0.40 | -0.45 | 0.05 | 0.096 | -0.32 | -0.37 | 0.06* | 0.019 |
| **Number of Students** | **20,528** | **20,324** | | | **20,346** | **20,230** | | |
| **Number of Schools** | **66** | **66** | | | **66** | **66** | | |
| **Year 2** | | | | | | | | |
| Grades 3–5 | -0.41 | -0.45 | 0.04 | 0.146 | -0.38 | -0.42 | 0.03 | 0.055 |
| Grades 6–8 | -0.34 | -0.39 | 0.05 | 0.101 | -0.31 | -0.34 | 0.02 | 0.281 |
| **Number of Students** | **20,252** | **20,457** | | | **20,032** | **20,359** | | |
| **Number of Schools** | **66** | **66** | | | **66** | **66** | | |

Source: Student administrative data.

Note: The difference between the treatment and control estimates may not equal the impact shown in the table due to rounding.

*Impact is statistically significant at the .05 level, two-tailed test.

## Supplemental Information for Systematic Reviews

Systematic reviews of evidence on the impacts of educational interventions often require specific types of information to evaluate the quality of a study. This section provides supplemental information that a systematic review would potentially need to assess the quality of the main impact findings reported in Chapter VI—specifically, findings about the impacts of pay-for-performance on educator effectiveness and student achievement in Cohort 1 schools.

### Cluster and School Attrition

Because this study was a randomized controlled trial, the extent of attrition from the original randomly assigned sample is the key factor determining the quality of the impact findings. As discussed in Appendix A, we randomly assigned clusters—either schools or groups of schools—to the treatment or control groups. We then made conclusions (or "inferences") about the impacts of pay-for-performance on schools, a subcluster unit. Therefore, the attrition rates of both clusters and schools are central to evaluating the evidence in Chapter VI.

Table F.18 shows the original number of clusters that we randomly assigned and the final number of clusters included in the analysis of each outcome. Among the original ("baseline") sample of clusters relevant to most outcomes, we assigned 48 clusters to the treatment group and 48 clusters to the control group. Some educator effectiveness outcomes were not applicable to particular districts because either the districts did not use those types of effectiveness measures or those measures were not based on a rating scale with a defined minimum and maximum value. Whenever an outcome was not applicable to a particular district, we excluded the treatment and control clusters in that district from the definition of the original, randomly assigned sample. For each outcome, the number of clusters in the final analysis sample differed from the original number of randomly assigned clusters due to cases in which (1) all schools in a cluster closed or dropped out of the study; (2) the study team dropped clusters that, for random assignment, had been paired with clusters that closed or dropped out; or (3) all schools in a cluster had missing data on the specified outcome.

School attrition (within clusters that remained in the study) also determines the quality of the impact findings because, for every outcome examined in Chapter VI, we sought to make conclusions about impacts on schools. As explained in Chapters I, II, and VI, pay-for-performance could affect the average educator effectiveness of schools in the study by either enabling schools to retain and recruit more effective educators or motivating educators to improve their performance. Impacts on average educator effectiveness in the study schools, reported in Tables VI.1 and VI.2, could reflect a combination of these influences. Likewise, as stated in Chapter VI, the study's findings on student achievement at the end of Year 1 captured the "the impact of pay-for-performance on schools' average student achievement after the first year of implementation," and the findings at the end of Year 2 captured the "cumulative impact on schools' average student achievement after two years of implementation" (see page 90). In Chapter II, we explained that these impacts on student achievement were "potentially reflecting changes in individual students' achievement and changes in the schools' student composition resulting from pay-for-performance" (pages 23 to 24). Therefore, for the outcomes examined in Chapter VI, the units for which we made inferences (schools) were not the same as the ultimate units of analysis (educators or students).

The final four columns of Table F.18 show the original number of schools at the time of random assignment and the final number of schools included in the analysis of each outcome. Both types of school counts are based only on the clusters that remained in the analysis for the specified outcome.

**Effect Sizes**

Table F.19 provides complete information needed for computing effect sizes. The adjusted mean outcomes, impacts, and *p*-values are identical to those reported in Chapter VI. The additional information in this table consists of the unadjusted standard deviations of the outcomes in the treatment and control groups.

**Table F.18. Cluster and School Attrition in the Analysis of the Impacts of Pay-for-Performance on Educator Effectiveness and Student Achievement, Cohort 1**

| | Original Number of Clusters that were Randomly Assigned | | Final Number of Clusters that Remained in the Analysis Sample | | Original Number of Schools in the Remaining Clusters | | Final Number of Schools that Remained in the Analysis Sample | |
|---|---|---|---|---|---|---|---|---|
| Outcome | Treatment | Control | Treatment | Control | Treatment | Control | Treatment | Control |
| **Outcomes Examined in Table VI.1** | | | | | | | | |
| School Achievement Growth Ratings, Year 1 | 44[a] | 44[a] | 41 | 41 | 62 | 62 | 62 | 62 |
| School Achievement Growth Ratings, Year 2 | 48 | 48 | 45 | 44 | 66 | 65 | 66 | 65 |
| Classroom Achievement Growth Ratings, Year 1 | 23[b] | 23[b] | 21 | 21 | 37 | 36 | 37 | 36 |
| Classroom Achievement Growth Ratings, Year 2 | 23[b] | 23[b] | 21 | 21 | 37 | 36 | 37 | 36 |
| **Outcomes Examined in Table VI.2** | | | | | | | | |
| Teachers' Classroom Observation Ratings, Year 1 | 48 | 48 | 45 | 45 | 66 | 66 | 66 | 66 |
| Teachers' Classroom Observation Ratings, Year 2 | 48 | 48 | 45 | 45 | 66 | 66 | 66 | 66 |
| Observation Ratings for Principals, Year 1 | 48 | 48 | 37 | 37 | 55 | 55 | 53 | 52 |
| Observation Ratings for Principals, Year 2 | 48 | 48 | 43 | 40 | 64 | 61 | 61 | 56 |
| **Outcomes Examined in Table VI.3** | | | | | | | | |
| Student Math Achievement, Year 1 | 48 | 48 | 45 | 45 | 66 | 66 | 66 | 66 |
| Student Math Achievement, Year 2 | 48 | 48 | 45 | 45 | 66 | 66 | 66 | 66 |
| Student Reading Achievement, Year 1 | 48 | 48 | 45 | 45 | 66 | 66 | 66 | 66 |
| Student Reading Achievement, Year 2 | 48 | 48 | 45 | 45 | 66 | 66 | 66 | 66 |

Source: Educator and student administrative data.

[a]Count excludes one district that did not structure school achievement growth ratings on a rating scale with a defined minimum and maximum value. Neither treatment nor control schools from this district are included in the count.

[b]Count excludes four districts that did not use classroom achievement growth to evaluate teachers. Neither treatment nor control schools from those four districts are included in the count.

**Table F.19. Detailed Statistics About the Impacts of Pay-for-Performance on Educator Effectiveness and Student Achievement After Years 1 and 2 (Points on 1-to-4 rating scale unless otherwise noted)**

| | Treatment Schools | | Control Schools | | | |
|---|---|---|---|---|---|---|
| Outcome | Adjusted Mean | Unadjusted Standard Deviation | Adjusted Mean | Unadjusted Standard Deviation | Impact | *p*-value |
| **Outcomes Examined in Table VI.1** | | | | | | |
| School Achievement Growth Ratings, Year 1 | 2.59 | 1.00 | 2.25 | 0.99 | 0.34* | 0.046 |
| School Achievement Growth Ratings, Year 2 | 2.46 | 0.96 | 2.21 | 0.89 | 0.25* | 0.047 |
| Classroom Achievement Growth Ratings, Year 1 | 2.26 | 0.96 | 2.08 | 0.95 | 0.18* | 0.033 |
| Classroom Achievement Growth Ratings, Year 2 | 2.20 | 0.99 | 2.16 | 1.04 | 0.04 | 0.459 |
| **Outcomes Examined in Table VI.2** | | | | | | |
| Teachers' Classroom Observation Ratings, Year 1 | 2.94 | 0.51 | 2.91 | 0.55 | 0.03 | 0.243 |
| Teachers' Classroom Observation Ratings, Year 2 | 3.02 | 0.48 | 2.97 | 0.52 | 0.05 | 0.070 |
| Observation Ratings for Principals, Year 1 | 3.08 | 0.60 | 3.18 | 0.60 | -0.10 | 0.197 |
| Observation Ratings for Principals, Year 2 | 3.16 | 0.68 | 3.03 | 0.71 | 0.13 | 0.184 |
| **Outcomes Examined in Table VI.3** | | | | | | |
| Student Math Achievement, Year 1 (student *z*-score units) | -0.43 | 0.93 | -0.45 | 0.93 | 0.02 | 0.335 |
| Student Math Achievement, Year 2 (student *z*-score units) | -0.39 | 0.92 | -0.43 | 0.92 | 0.04 | 0.068 |
| Student Reading Achievement, Year 1 (student *z*-score units) | -0.37 | 0.95 | -0.40 | 0.96 | 0.03* | 0.040 |
| Student Reading Achievement, Year 2 (student *z*-score units) | -0.36 | 0.95 | -0.39 | 0.95 | 0.03* | 0.026 |

Source: Educator and student administrative data.

Note: Means were adjusted by the regression model described in Appendix B. Unadjusted standard deviations were the standard deviations across schools for school achievement growth outcomes; across teachers for teachers' performance rating outcomes; across principals for principals' performance rating outcomes; and across students for student achievement outcomes.

# APPENDIX G

# SUPPLEMENTAL FINDINGS ON RELATIONSHIPS BETWEEN TIF PROGRAM CHARACTERISTICS AND THE IMPACTS OF PAY-FOR-PERFORMANCE FOR CHAPTER VI

THIS PAGE IS INTENTIONALLY BLANK

This appendix supplements the information presented in Chapter VI on the relationships between districts' TIF program and implementation characteristics and the impacts of pay-for-performance in Year 2. In this appendix, we provide (1) the rationale for choosing the characteristics we examined, (2) information on how we characterized districts into subgroups based on their characteristics, (3) analyses that examine the association between district program characteristics and impacts on math and reading achievement, and (4) analyses of the relationship between program characteristics and impacts on teacher retention between Years 1 and 3.

As discussed in Chapter II, evaluation districts were classified into two cohorts—Cohort 1 and Cohort 2—according to the year in which we randomly assigned their schools to a treatment group or a control group. The 10 districts whose schools were randomly assigned in spring and summer 2011 were classified as Cohort 1. Three additional districts, whose schools were randomly assigned in spring and summer 2012, were classified as Cohort 2. Cohort 1 completed two years of implementation during the period covered by this report, 2011–2012 and 2012–2013, referred to as Years 1 and 2. Cohort 2 districts completed only one year of implementation, 2012–2013, referred to as Year 1 for this cohort.

The information and analyses in this appendix pertain to Cohort 1 only.

## Characteristics Examined

We examined the relationships between four key characteristics and the impacts of pay-for-performance. Two characteristics—the use of classroom achievement growth to measure teacher effectiveness and the degree of differentiation in performance bonuses awarded—pertain to how the programs were designed. The other two characteristics—teachers' understanding of their eligibility for performance bonuses and the timing of bonus notification and award—relate to how the programs were implemented. We selected these four characteristics because of their potential to motivate teachers to change their behavior in response to pay-for-performance bonuses, which may, in turn, affect student achievement. These characteristics also varied across districts.[82]

### Districts' Use of Classroom Achievement Growth

Districts' use of classroom achievement growth to measure teacher effectiveness was the critical factor in shaping whether teachers' pay-for-performance bonuses were primarily determined by their own performance or that of a larger group of teachers. In the six districts that used these measures, most of the potential bonus amount that teachers could earn was based on their own performance; in the remaining districts, most of the bonus amount was based on the performance of their school or instructional team (see Chapter IV, Figure IV.5). An emphasis on individual, rather than group, incentives could have a positive, negative, or no association with impacts on student achievement. On the one hand, teachers might be more motivated to respond to individual incentives because they have more control over achievement growth in their classrooms than in their team or school. On the other

[82] We considered two other characteristics: the size of the performance bonus and the amount of TIF-required professional development that teachers received. However, districts with larger maximum bonuses were, in general, the same districts that had larger amounts of differentiation in bonuses, so we only report on the association between the amount of differentiation and impacts. There was little variation across districts in the average amount of pay-for-performance bonuses (see Figure IV.3), as well as in the amount of TIF-required professional development (in 9 of 10 districts, treatment teachers reported an average of five or fewer hours of TIF-required professional development). Therefore, we did not examine the association between these characteristics and impacts.

hand, individual incentives may involve comparing teachers with each other and could harm teacher collaboration.

### Amount of Differentiation in Teachers' Pay-for-Performance Bonuses

The amount of differentiation in teachers' pay-for-performance bonuses determined how much more compensation a teacher could earn by performing well than by performing poorly. If teachers are motivated by the ability to earn a larger bonus, then more differentiation in bonus amounts could increase teacher productivity and improve student achievement. On the other hand, for those who believe that teachers should be paid similarly (or based on tenure), pay-for-performance with large differences in payouts among teachers may lower satisfaction and have a negative impact on teachers' productivity. We classified 3 of 10 districts as having high amounts of differentiation in their pay-for-performance bonuses for teachers.

### Teachers' Understanding of Their Eligibility for Pay-for-Performance Bonuses

Teachers must understand they are eligible for pay-for-performance bonuses for those bonuses to affect their decisions and behavior. Therefore, we created subgroups based on the level of teachers' understanding of their eligibility for pay-for-performance bonuses. Because treatment teachers were supposed to be eligible for performance bonuses and control teachers were not, we grouped districts by their differences in perceived eligibility between treatment and control teachers. We classified 4 of 10 districts as having high levels of teacher understanding (at least a 50 percentage point difference between treatment and control schools in the percentage of teachers who believed they were eligible for a pay-for-performance bonus in Year 2).

### Timing of Bonus Notification and Award

We also created subgroups based on when teachers were notified and awarded performance bonuses. When teachers learn about the performance bonuses that they or their colleagues have earned, they may become more aware of their eligibility, the criteria for earning a bonus, and the size of the bonuses. However, teachers need time to translate this new knowledge into actions that can improve their effectiveness. We classified 3 of the 10 districts as having early timing of bonus notification and award (districts that notified and awarded at least one component of their Year 1 bonuses by the August after the 2011–2012 school year).

## Categorizing Districts into Subgroups Based on Program and Implementation Characteristics

This section provides details on how Cohort 1 districts were categorized into two subgroups based on each of the program characteristics examined in Chapter VI. For each characteristic, we categorized districts into two subgroups that differed according to the presence or absence of the characteristic, or according to the extent (high or low) of that characteristic.

Table G.1 describes the program and implementation characteristics used in the subgroup analyses and how districts were categorized by the characteristic. The final column in Table G.1 indicates the number of districts that met the subgroup definition described in the table. The remaining districts (out of the 10 Cohort 1 districts) made up its comparison subgroup.

**Table G.1. Program and Implementation Characteristics Used for Subgroup Analysis**

| Characteristic | Reason for Examining This Characteristic | Subgroup Definition | Number of Districts in Subgroup |
|---|---|---|---|
| Awarded bonuses based on classroom achievement growth[a] | Measure increases emphasis on individual over group performance, which may enhance teachers' control over their own ratings but discourage collaboration. | Districts that used classroom achievement growth measures in determining performance bonuses were assigned a 1. | 6 |
| Degree of differentiation in performance bonuses[b] | Differentiated bonuses increase the monetary gain from being a high performer which may provide greater motivation to improve teacher productivity. | Districts where standard deviation of performance bonuses in Year 1 was at least 5 percent of average teacher salary were assigned a 1. | 3 |
| Teachers' understanding of their eligibility for pay-for-performance bonuses[c] | Understanding of eligibility is necessary for bonuses to affect behavior. | Difference between the percentages of teachers in treatment and control schools who believed they were eligible for pay-for-performance bonuses exceeded 50 percentage points. | 4 |
| Timing of bonus notification and award[d] | Early notification and award of bonuses from one year allow more time for teachers to change schools, adjust their understanding, and revise their teaching practices for the next year. | Districts carried out notification and bonus awards of at least one component of Year 1 bonuses no later than the August after Year 1. | 3 |

[a]Based on district interviews.

[b]Based on educator administrative data from Year 1.

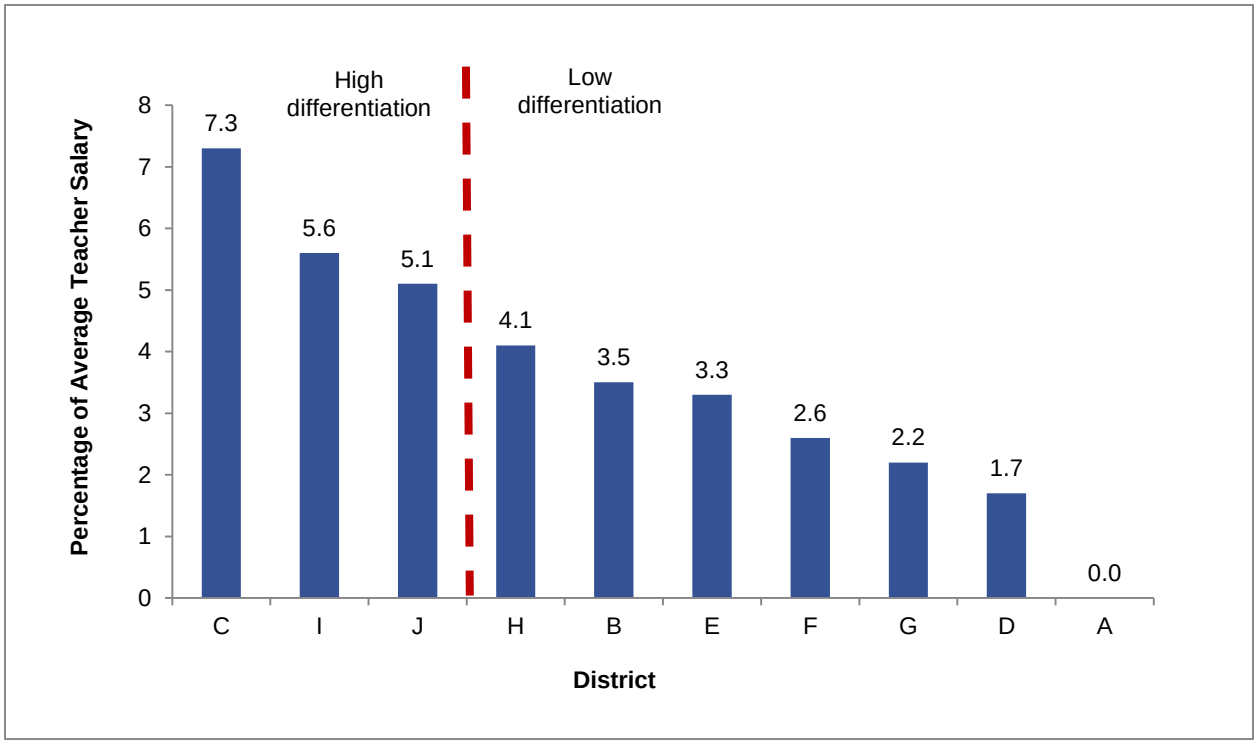
[c]Based on teacher survey, spring 2013.

[d]Based on district interviews. One district that failed to award bonuses from Year 1 is included in the subgroup that did not provide notification and awarding of bonuses by the August after Year 1.

Chapter IV, Table IV.2 shows the percentage of Cohort 1 districts that used different types of student achievement growth measures to evaluate teachers, including those that evaluated teachers based on the achievement growth of the teachers' own students (classroom achievement growth). In addition, as explained in Chapter IV, of the nine districts that paid out any bonuses, three reported notifying and paying teachers before the start of the 2012–2013 school year. The remaining six districts reported notifying and paying teachers between October and December 2012.

Although the grant notice provided an example of a sufficiently differentiated bonus program as one in which a teacher could earn a bonus at least three times the average amount, our analysis used a more formal measure of differentiation. For our subgroup analysis, we used a measure, the standard deviation, which captured how extensively below- and above-average bonuses differed in dollar value from the average bonus. Our analysis classified districts as having high amounts of differentiation if the standard deviation of bonus amounts within the district in Year 1 exceeded 5 percent of average teacher salary. Districts that met the grant notice's example of differentiation but had very small bonuses would not be classified as a district with high differentiation of bonuses by our measure because the dollar value of the differences in bonus amounts between teachers would still be small. Figure G.1 shows the groupings of districts for the degree of differentiation in performance bonuses.

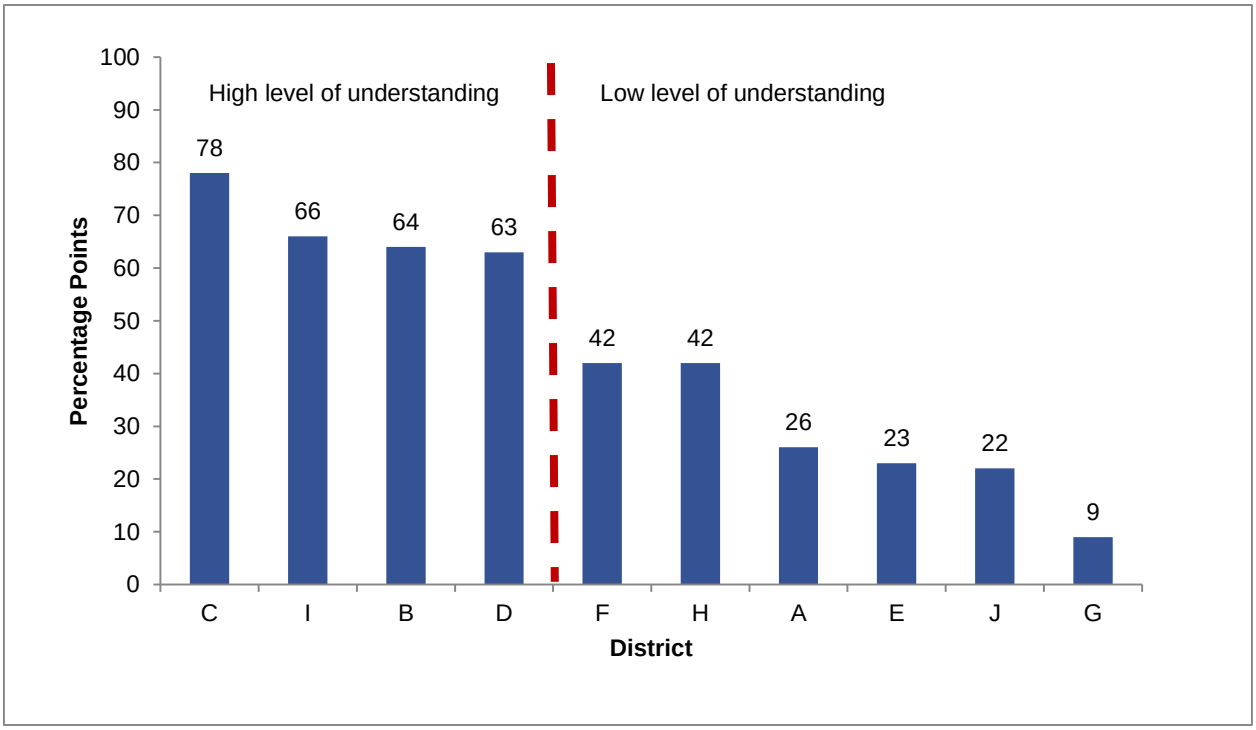
**Figure G.1. Standard Deviation of Performance Bonuses in Year 1 as a Percentage of Average Teacher Salary, Cohort 1**




Source: Educator administrative data, Year 1 (N = 2,189 teachers) and district interviews.

Figure G.2 illustrates the groupings of districts for the extent of teachers' understanding of their eligibility for pay-for-performance. Districts classified as having high levels of teacher understanding were those that had at least a 50 percentage point difference between treatment and control schools in the percentage of teachers who believed they were eligible for a pay-for-performance bonus in Year 2.

**Figure G.2. Difference Between the Percentages of Teachers in Treatment and Control Schools Who Believed They Were Eligible for Performance Bonuses in Year 2, Cohort 1**




Source: Educator survey data, Year 1 (N = 893 teachers).

## Association Between Program Characteristics and Impacts on Math and Reading Achievement

### Subgroup Analyses

For each pair of subgroups that differed on a particular characteristic, we estimated the impacts of pay-for-performance on student achievement in Year 2 within the two subgroups and ascertained whether the impacts differed between the subgroups. A statistically significant difference in impacts between the two subgroups would represent an association between the characteristic and impacts. As discussed in Chapters II and VI, we expressed achievement outcomes as $z$-scores based on statewide means and standard deviations of scores in each grade.

There was little evidence that key TIF program or implementation characteristics could explain differences across districts in the impacts of pay-for-performance on student achievement. Of the four characteristics we examined, only one had a statistically significant relationship with student achievement. Higher differentiation in bonuses had a negative association with impacts on math achievement in Year 2 and no association with impacts on reading achievement (Table G.2). In math, the impact in the three districts with high amounts of differentiation was lower by 0.08 standard deviations compared to the impact in the remaining districts.

**Table G.2. Differences in Year 2 Impacts on Student Achievement Between Subgroups Based on Districts' Program Characteristics**

| | Math | | Reading | |
|---|---|---|---|---|
| Subgroup of Districts with… | Difference in Impacts Between Specified Subgroup and Remaining Districts (student *z*-score units) | *p*-value | Difference in Impacts Between Specified Subgroup and Remaining Districts (student *z*-score units) | *p*-value |
| Classroom Achievement Growth Factored into Pay-for-Performance Bonuses | 0.06 | 0.240 | 0.03 | 0.311 |
| High Level of Differentiation in Pay-for-Performance Bonuses | -0.08* | 0.027 | -0.01 | 0.601 |
| High Level of Teacher Understanding of Pay-for-Performance Eligibility | -0.01 | 0.808 | 0.00 | 0.997 |
| Early Notification and Award of Bonuses | 0.02 | 0.733 | 0.01 | 0.646 |
| **Number of Students** | **40,709** | | **40,391** | |
| **Number of Schools** | **132** | | **132** | |

Source: Student administrative data.

*Difference is statistically significant at the .05 level, two-tailed test.

None of the other characteristics that we examined—the use of classroom achievement growth measures, teachers' understanding of pay-for-performance eligibility, or the timing of the Year 1 bonus notification and award—had a statistically significant relationship with impacts on student achievement in math or reading in Year 2.

## Sensitivity Analyses

Instead of using the numeric values of program characteristics to categorize districts into two subgroups, Table G.3 shows the results when we directly examined whether these numeric values were associated with impacts on math and reading achievement in Year 2. None of these analyses found any relationship between program characteristics and impacts on math or reading achievement in Year 2.

**Table G.3. Association Between Continuous Measures of Program Characteristics and Impacts on Student Achievement in Year 2, Cohort 1**

| | Association with Impacts in Math (student z-score units) | | Association with Impact in Reading (student z-score units) | |
|---|---|---|---|---|
| Program Characteristic | Coefficient | *p*-value | Coefficient | *p*-value |
| Standard Deviation of Performance Bonuses in Year 1 as a Percentage of Average Teacher Salary | -0.010 | 0.294 | 0.003 | 0.574 |
| Difference Between the Percentages of Teachers in Treatment and Control Schools Who Believed They Were Eligible for Performance Bonuses in Year 2 | -0.001 | 0.564 | 0.000 | 0.971 |
| Number of Months Since May 2012 When District Paid Out Year 1 Performance Bonuses[a] | 0.003 | 0.838 | 0.002 | 0.689 |
| **Number of Students—Range[b]** | **36,271-40,709** | | **36,006-40,391** | |
| **Number of Schools—Range[b]** | **114-132** | | **114-132** | |

Source: Student administrative data.

[a]Estimates exclude the district that failed to award bonuses from Year 1.

[b]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

THIS PAGE IS INTENTIONALLY BLANK

THIS PAGE IS INTENTIONALLY BLANK


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA