
Evaluation of the Teacher Incentive Fund: Implementation and Early Impacts of Pay-for-Performance After One Year

September 2014



U.S. Department of Education

THIS PAGE IS INTENTIONALLY BLANK

Evaluation of the Teacher Incentive Fund: Implementation and Early Impacts of Pay-for-Performance After One Year

September 2014

Jeffrey Max
Jill Constantine
Alison Wellington
Kristin Hallgren
Steven Glazerman
Hanley Chiang
Cecilia Speroni
Mathematica Policy Research

Elizabeth Warner
Project Officer
Institute of Education Sciences

NCEE 2014-4019
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

Sue Betka

Acting Director

National Center for Education Evaluation and Regional Assistance

Ruth Curran Neild

Commissioner

September 2014

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.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be:

Max, Jeffrey, Jill Constantine, Alison Wellington, Kristin Hallgren, Steven Glazerman, Hanley Chiang, Cecilia Speroni. (2014). *Evaluation of the Teacher Incentive Fund: Implementation and Early Impacts of Pay-for-Performance After One Year* (NCEE 2014-4019). 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>.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

ACKNOWLEDGMENTS

This study would not have been possible without the contributions of many individuals. We are deeply grateful to the many TIF administrators, teachers, principals, district leaders, and central office staff whose hard work and patience made this research possible. The technical assistance team worked closely and thoughtfully with district TIF staff to support program implementation. We are grateful to Duncan Chaplin for his leadership and 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 for all of their efforts on behalf of districts and the study.

This report also relies heavily on district, teacher, and principal surveys. At Mathematica, Sheila Heaviside and Annette Luyegu led all aspects of survey administration. We also thank Jacqueline Agufa and the team who collected information on teacher placements from all TIF districts, and Kathy Shepperson, who made sure we didn't lose track of anyone.

We thank Chris Jones for his patience and flexibility in jumping in on any task needed on the study. We also thank Raul Torres Aragon and Katharine Lindquist who provided expert programming. Melissa Clark provided helpful comments on earlier versions of the report. A technical working group (TWG) provided useful input on study design and findings. TWG members included David Heistad, Jim Kemple, Dan McCaffrey, Dick Murnane, Anthony Milanowski, Jeffrey Smith, and Jacob Vigdor. Cindy George and John Kennedy oversaw the editing of the report. Jill Miller prepared the report for publication with great care and patience.

THIS PAGE IS INTENTIONALLY BLANK

CONTENTS

EXECUTIVE SUMMARY.....	xv
I INTRODUCTION.....	1
II STUDY SAMPLE, DESIGN, DATA, AND METHODS.....	13
III TIF DISTRICTS AND THEIR PROGRAMS.....	25
IV IMPLEMENTATION OF TIF IN THE EVALUATION DISTRICTS.....	37
V INTERMEDIATE IMPACTS ON EDUCATORS' ATTITUDES AND BEHAVIORS	69
REFERENCES.....	81
APPENDIX A: SUPPLEMENTARY INFORMATION ON STUDY SAMPLE AND DESIGN.....	A.1
APPENDIX B: SURVEY RESPONSE RATES AND CHARACTERISTICS OF RESPONDENTS	B.1
APPENDIX C: ANALYTIC METHODS AND SENSITIVITY ANALYSES.....	C.1
APPENDIX D: SUPPLEMENTAL FINDINGS ON TIF DESIGN AND IMPLEMENTATION FOR CHAPTERS III AND IV.....	D.1
APPENDIX E: SUPPLEMENTARY FINDINGS FOR CHAPTER V.....	E.1

THIS PAGE IS INTENTIONALLY BLANK

TABLES

ES.1	Performance Measures Used to Evaluate Teachers and Principals, as Reported by Educators and District Representatives	xxv
ES.2	Teachers' Satisfaction with Performance Measures, Professional Opportunities, School Environment, and TIF Program	xxvii
II.1	TIF Grants Awarded, Grantees Implementing TIF in 2011–2012 School Year.....	14
II.2	Comparison of TIF Evaluation Districts and Non-Evaluation Districts (Percentages Unless Otherwise Noted)	16
II.3	Data Sources for First TIF Report	18
III.1	Percentage of Districts Using Measures of Student Achievement to Evaluate Teachers	26
III.2	Classroom Observations to Evaluate Teachers in TIF Districts	26
III.3	District Report About Principal Evaluation Measures (Percentages).....	27
III.4	Additional Pay Opportunities for Teachers and Principals.....	32
III.5	Planned Focus of Professional Development (Percentages)	33
III.6	Implementation of TIF Program Requirements (Percentages)	35
III.7	Type of Educator Involvement in TIF Program Development (Percentages).....	35
III.8	Reasons Districts Reported for Revising Their Proposed TIF Program (Percentages)	36
IV.1	Percentage of Districts Using Student Achievement and Observation Measures for Teachers and Principals, by Evaluation Participation Status	39
IV.2	Additional Pay Opportunities for Teachers, Comparison of TIF Evaluation and Non-Evaluation Districts	45
IV.3	Percentage of Districts Reporting Professional Development Activities for Teachers Planned Under TIF, by Evaluation Participation Status	45
IV.4	Percentage of Districts Reporting Implementation of TIF Program Requirements, by Evaluation Participation Status.....	47

IV.5	Educator Involvement in TIF Program Development (percentages), by Evaluation Participation Status	47
IV.6	Reasons Districts Reported for Revising Their Proposed TIF Programs, by Evaluation Participation Status (percentages)	48
IV.7	Performance Measures Used to Evaluate Teachers, as Reported by Educators and District Representatives (percentages unless otherwise noted).....	55
IV.8	Performance Measures Used to Evaluate Principals, as Reported by Principals and District Representatives	55
IV.9	Additional Teacher Pay for Extra Roles and Responsibilities, as Reported by Educators and District Representatives	56
IV.10	Teachers' Reports of the Professional Development They Received (percentages)	67
IV.11	Teachers' Reports of Hours Spent in Professional Development Activities (averages).....	67
V.1	Teachers' Satisfaction with Performance Measures, Professional Opportunities, and School Environment (Percentages Who Are "Somewhat" or "Very" Satisfied)	70
V.2	Difference Between Teachers in Treatment and Control Schools on Selected Teacher-Satisfaction Measures, by Subgroup (Percentage Points).....	72
V.3	Principal Satisfaction with Professional Opportunities and School Environment (Percentages Who Were "Somewhat" or "Very" Satisfied)	72
V.4	Teachers' Attitudes Toward TIF Program (Percentages Who "Agreed" or "Strongly Agreed").....	73
V.5	Principals' Attitudes Toward TIF Program (Percentages Who "Agreed" or "Strongly Agreed").....	74
V.6	Teachers' Average Time Spent on School-Related Activities in the Most Recent Full Week (Averages, in Hours)	75
V.7	Incentives Used to Recruit Teachers (Percentages Who Reported They Were "Always" or "Often" Used)	77
V.8	Teaching Vacancies and Hiring Experiences (Averages Unless Otherwise Noted)	77

V.9	Criteria Used to Assign Teachers to Grade Levels or Subject Areas (Percentages Who Report They Are “Always” or “Often” Used).....	78
V.10	Influence of TIF Program on Educators' School Preference (Percentages).....	79
V.11	Influence of TIF Program on Teachers' School Preference for Next Year (Percentages).....	79
A.1	Comparison of TIF Districts to All U.S. Districts (Percentages Unless Otherwise Noted)	A.3
A.2	Average Baseline Characteristics of Treatment and Control Schools	A.5
B.1	District Survey Response Rates Overall and by Evaluation Status.....	B.3
B.2	District Characteristics by District's Response Status (Percentages Unless Otherwise Noted)	B.4
B.3	Teacher Respondents by Teaching Assignment and Treatment Status	B.5
B.4	Teacher and Principal Survey Response Rates Overall and by Treatment Status.....	B.6
B.5	Survey Response Rates, Teacher Survey (Percentages).....	B.7
B.6	Survey Response Rates, Principal Survey (Percentages)	B.8
B.7	School and Student Characteristics of Teacher Survey Respondents and Full Teacher Sample (Percentages).....	B.9
B.8	School and Student Characteristics of Principal Survey Respondents and Full Principal Sample (Percentages).....	B.10
B.9	District Reported Characteristics by District's Response Status on the Survey Question about the Distribution of Pay-for-Performance Bonuses for Teachers (Percentages of Districts)	B.11
C.1	Interclass Correlation in Multidistrict Grantees	C.3
D.1	Percent of Districts Using Additional Evidence To Evaluate Teachers and Principals.....	D.3
D.2	Districts' Past Experience Providing Pay-for-Performance Bonuses or Opportunities for Additional Pay (Percentages)	D.5
D.3	Additional Pay Opportunities for Teachers, Compared with Pay-for-Performance Bonus.....	D.5

D.4	Districts' Reports of Teacher Evaluation Measures (Percentage Unless Otherwise Indicated)	D.7
D.5	District Report About Additional Evidence Used for Principal Evaluations (Percentages)	D.7
D.6	Comparison of Additional Pay Opportunities and Pay-for-Performance Bonuses, by Evaluation Status	D.9
D.7	Additional Evidence Used to Evaluate Teacher Performance, as Reported by Educators and District Representatives	D.10
D.8	Additional Evidence Used to Evaluate Principal Performance, as Reported by Principals and District Representatives.....	D.10
D.9	Types of Additional Pay for Teachers, as Reported by Educators and District Representatives	D.11
D.10	Percentage of Teachers Receiving Professional Development, as Reported by Educators and District Representatives	D.12
D.11	Teacher Performance Measures Used, as Reported by Teachers in Tested and Nontested Grades and Subjects.....	D.13
D.12	Teachers' Eligibility for Pay-for-Performance Bonuses, as Reported by Teachers: Subgroup Analyses	D.14
E.1	Impacts on Satisfaction Using Alternative-Outcome Definitions.....	E.4
E.2	Teachers' Attitudes Toward TIF Program Using Alternative-Outcome Specifications	E.5
E.3	Principals' Attitudes Toward TIF Program Using Alternative-Outcome Specifications.....	E.6
E.4	Incentives Used to Recruit Teachers Using Alternative-Outcome Specifications	E.7
E.5	Criteria Used for Teacher Assignments to Grade Levels or Subject Areas Using Alternative-Outcome Specifications	E.8
E.6	Teacher and Principal Satisfaction with Professional Opportunities and School Environment Using Alternative-Model Specifications (Percentages Who Are "Somewhat" or "Very" Satisfied).....	E.10
E.7	Teachers' Attitudes Toward TIF Program Using Alternative-Model Specifications (Percentages Who "Agree" or "Strongly Agree")	E.11

E.8	Incentives Used to Recruit Teachers Using Alternative-Model Specifications (Percentages Who Report They Are “Always” or “Often” Used)	E.12
E.9	Criteria Principals Used for Assigning Teachers to Grade Levels or Subject Areas Using Alternative-Model Specifications (Percentages Who Report They Are “Always” or “Often” Used)	E.13
E.10	Influence of TIF Program on Educators' School Preference (Percentages).....	E.14
E.11	Demographic Characteristics, Educational Background, and Certification of Teachers (Percentages)	E.15
E.12	Teacher Satisfaction by Subgroup (Percentages That Are “Somewhat” or “Very” Satisfied)	E.17
E.13	Teachers' Attitudes Toward TIF Program by Subgroup (Percentages Who “Agree” or “Strongly Agree”)	E.19
E.14	Demographic Characteristics of Teachers and Principals (Percentages Unless Otherwise Noted)	E.22
E.15	Educational Background and Certification of Teachers (Percentages).....	E.23
E.16	Work Experience of Teachers and Principals (Averages Unless Otherwise Noted)	E.24
E.17	Educational Background and Certification of Teachers, by Teachers' Grade Level (Percentages)	E.25

THIS PAGE IS INTENTIONALLY BLANK

FIGURES

ES.1	Random Assignment Evaluation Design	xix
ES.2	Expected Pay-for-Performance Bonuses for Teachers and Principals in Evaluation Districts, Averages Across Districts	xxiii
ES.3	Relative Weight of Each Type of Performance Measure Used for Pay-for-Performance Bonuses in TIF Evaluation Districts.....	xxiv
ES.4	Maximum Possible Size of Pay-for-Performance Bonuses for Treatment Teachers and Principals, as Reported by Respective Educators and Districts	xxv
I.1	Logic Model.....	9
II.1	Characteristics of TIF Districts Compared with All U.S. Districts.....	15
II.2	Random Assignment Design.....	17
III.1	Average, Minimum, and Maximum Expected Pay-for-Performance Bonuses for Teachers and Principals.....	29
III.2	Expected Distribution of Teacher Pay-for-Performance Bonuses in Tested Grades and Subjects.....	31
III.3	Expected Distribution of Principal Pay-for-Performance Bonuses.....	31
IV.1	Expected Pay-for-Performance Bonuses for Teachers in Evaluation and Non-Evaluation Districts, Averages Across Districts	41
IV.2	Expected Pay-for-Performance Bonuses for Principals in Evaluation and Non-Evaluation Districts, Averages Across Districts	42
IV.3	Distribution of Expected Pay-for-Performance Bonuses for Teachers in Tested Grades and Subjects, by Evaluation Participation Status.....	43
IV.4	Distribution of Expected Pay-for-Performance Bonuses for Principals, by Evaluation Participation Status	43
IV.5	Relative Weight of Each Type of Performance Measure Used for Pay-for-Performance Bonuses in TIF Evaluation Districts.....	51
IV.6	Relative Weight of Each Type of Measure Used for Performance Bonuses for Teachers in Tested Grades and Subjects in Evaluation Districts	52
IV.7	Teachers' Pay-for-Performance Bonus Eligibility, as Reported by Teachers and Principals.....	57

IV.8	Maximum Possible Size of Pay-for-Performance Bonuses for Teachers, as Reported by Teachers and Principals.....	59
IV.9	Teachers’ Automatic Bonus Eligibility, as Reported by Teachers and Principals.....	60
IV.10	Principals’ Reports of Their Own Eligibility for Pay-for-Performance Bonuses	61
IV.11	Principals’ Reports of Their Own Eligibility for Automatic Bonuses	62
IV.12	Principals’ Reports of the Maximum Possible Size of Their Pay-for-Performance Bonuses.....	62
IV.13	Teachers’ Reports of the Measures Used to Evaluate Their Own Performance.....	64
IV.14	Principals’ Reports of the Measures Used to Evaluate Their Own Performance.....	65
IV.15	Teachers’ Reports of Whether Teachers in Their Schools Were Eligible for Additional Pay Opportunities	66
D.1	Pay-for-Performance Bonuses for Teachers in Tested Grades and Subjects, by Districts	D.4
D.2	Maximum Possible Size of Pay-for-Performance Bonuses for Teachers, as Reported by Teachers and Principals Who Provided Nonmissing Responses.....	D.15
D.3	Maximum Possible Size of Pay-for-Performance Bonuses for Principals, as Reported by Principals Who Provided Nonmissing Responses	D.16

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. The TIF grants have two goals:

- Reform compensation systems to reward educators for improving students' achievement
- Increase the number of high-performing teachers in high-need schools and hard-to-staff subjects

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 first of four planned reports from a multiyear study focusing on the TIF grants awarded in 2010.¹ It examines grantees' implementation experiences and intermediate educator outcomes near the end of the first year of program implementation, before the first pay-for-performance payouts to teachers and principals. Future reports will address the impacts of such payouts on student achievement, educator mobility, and changes in educators' job satisfaction and attitudes toward their TIF programs.

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 affect the retention and recruitment of educators and, ultimately, student achievement.

This report describes programs implemented during the 2011–2012 school year by the 2010 TIF grantees. The main findings for all TIF districts include the following:

- **Fewer than half of districts reported implementing all required components of the TIF program, evidence that full implementation is a challenge.** Although 85 percent of TIF districts reported implementing at least three of the four required components for teachers, slightly fewer than half (46 percent) reported implementing all four.²
- **Consistent with the TIF grant goals, grantees expected pay-for-performance bonuses to be somewhat substantial and differentiated. However, districts expected most educators would receive a bonus, suggesting that the award criteria were not consistent with TIF guidance for challenging pay-for-performance bonuses.** TIF districts expected to award an average pay-for-performance

¹ TIF grants are often 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 According to the original TIF notice, grantees could not use TIF program funds for incentive payments until they had implemented a performance-based compensation system that included all of the required components. Although most grantees used the 2010–2011 school year as a planning year, once grantees began implementation they were expected to implement all of the required components.

bonus of about 4 percent of the average U.S. educators' salary. The maximum bonus expected by TIF districts was twice as large as the average bonus for teachers and 50 percent larger than the average bonus for principals. Districts also expected to award a pay-for-performance bonus to more than 90 percent of eligible teachers and principals.

The report also provides detailed findings on implementation and the effect of pay-for-performance bonuses on educators for a subset of 2010 TIF grantees. These grantees include 10 districts participating in a random assignment study of the pay-for-performance component of the TIF program during the 2011–2012 school year. The key findings for the 10 evaluation districts include the following:

- **Many educators misunderstood the performance measures and the pay-for-performance bonuses used for TIF.** For example, the measures that educators indicated were used to evaluate their performance sometimes differed from those reported by districts. In addition, more than half of teachers did not know they were eligible for pay-for-performance bonuses, and teachers reported a maximum pay-for-performance bonus that was lower than the amount reported by districts.
- **Most teachers and principals are satisfied with their professional opportunities, school environment, and the TIF program.** About two-thirds of teachers were satisfied with their jobs overall and were glad to be participating in the TIF program.
- **Educators in schools that offered pay-for-performance bonuses tended to be less satisfied than those in schools that did not offer such bonuses.** For example, fewer teachers in schools that offered bonuses were satisfied with the opportunities for professional advancement (68 versus 76 percent) and school morale (48 versus 55 percent). However, more teachers in schools offering pay-for-performance bonuses were satisfied with the opportunity to earn additional pay (64 versus 59 percent).

TIF Grants and Requirements

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 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 program components in their performance-based compensation systems. This study focuses most heavily on one of those requirements: the impact of pay-for-performance bonuses.

Required Program Components of the Performance-Based Compensation Systems

The four required TIF program components are:

1. **Measures of educator effectiveness.** Grantees were required to use a measure of effectiveness for teachers and principals that included 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 were designed to

incentivize educators and reward them for being effective in their classrooms and schools. They had to be substantial, challenging, differentiated, 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 had to support teachers and principals in their performance-improvement efforts. Support included providing information on the measures on which educators would be evaluated and more targeted professional development based on an educator's actual performance on the effectiveness measures.

In addition to these required components, grantees could include another program component—offering incentives to recruit and retain effective educators in hard-to-staff subjects within high-need schools.

The TIF Grant Competition

The 2010 TIF grant application notice differed from the other rounds of the TIF grants in an important way: it included a main competition and an evaluation competition.³ Applicants had to apply for one or the other. 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 an evaluation to measure the impact of pay-for-performance bonuses on educators' and students' outcomes.

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 an impact evaluation of their TIF grants. They had to allow schools within a district to be 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 of the performance-based compensation system *except* pay-for-performance bonuses (the control group). Evaluation grantees also had to include at least eight elementary or middle schools in the evaluation and cooperate with additional data collection such as surveys of teachers and principals.

Another key difference between the main and evaluation grant requirements 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 payout worth 5 percent of the average educator's salary), differentiated (at least some educators could expect to receive a payout worth three times the average payout), and challenging to earn (only those who perform significantly better than average would receive bonuses). Although applicants for evaluation grants had discretion over the proposed structure of the pay-for-performance bonus, these examples provided additional guidance to applicants and could have influenced the design of their performance-based compensation systems.

³ The American Recovery and Reinvestment Act partially funded the 2010 TIF grants and also mandated a national evaluation of TIF.

The TIF Study

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

The study will address five research questions:

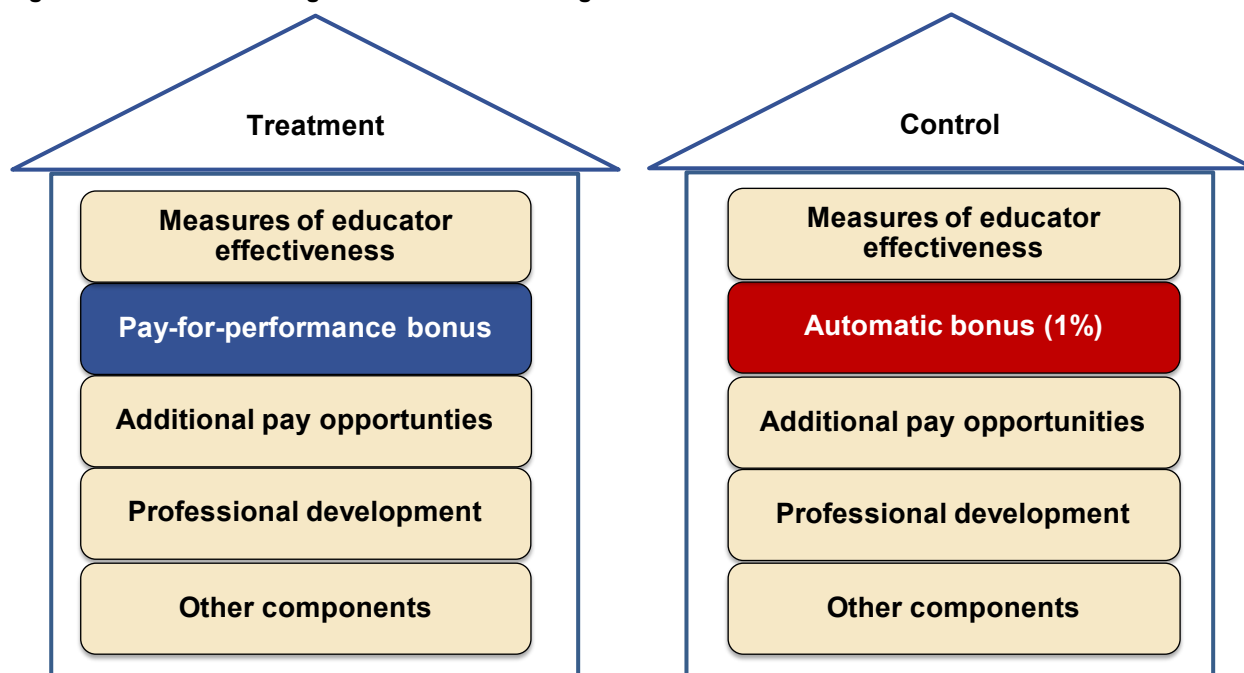
1. What are the characteristics of all TIF grantee districts and their performance-based compensation systems? What implementation experiences and challenges did TIF districts encounter?
2. How do teachers and principals in schools that did or did not offer pay-for-performance bonuses compare on key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

This study includes information on implementation of TIF for all 2010 grantees (question 1) and more in-depth implementation and impact information from a subset of 12 districts selected through the evaluation competition (questions 2 through 5). In this first report, the study team focuses on early implementation of the TIF grants (questions 1 and 2), specifically, the features of districts' performance-based compensation systems, the structure of the pay-for-performance bonuses, and educators' understanding of their districts' programs. In addition, for evaluation grantees, the study team examines the impact of pay-for-performance bonuses on intermediate outcomes related to educators' attitudes, productivity, recruitment, and retention near the end of the first year of implementation.

Study Design

In addition to an implementation analysis conducted for all 2010 TIF grantees, the study uses an experimental study design for districts that received TIF funding through evaluation grants. As shown in Figure ES.1, schools within the 12 evaluation districts were assigned randomly—that is, completely by chance—to treatment and control groups. 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 will measure the impact of pay-for-performance bonuses that are implemented within the context of broader performance-based compensation systems. The study is not designed to measure the impact of implementing a TIF grant or the multiple components of a performance-based compensation system.

Figure ES.1. Random Assignment Evaluation Design



Teachers and principals in treatment schools were eligible to earn a pay-for-performance bonus; teachers and principals in control schools were eligible to receive an automatic bonus worth approximately 1 percent of their salary. The TIF grant notice required the 1 percent bonus in control schools. The 1 percent bonus ensured that all educators in evaluation schools received some benefit from participating in the study, either a pay-for-performance bonus or the automatic bonus. Therefore, the impact of pay-for-performance estimated in this study is based on two potential effects (i) bonuses in treatment school were differentiated based on educator performance, and (ii) bonuses in treatment schools were a little larger on average, than in control schools. The random assignment process created two groups that, on average, were initially similar in terms of student achievement, school type, enrollment, school location, student race and ethnicity, and student socioeconomic status. This study design ensures that inferences about the effect of pay-for-performance bonuses are based solely on the offer of the bonuses and not on other characteristics of districts, schools, or educators.

Data Sources

Data for this report came from multiple sources. The sources enabled us to examine implementation broadly in all TIF districts and to report on more detailed aspects of implementation in the evaluation districts, including the experiences of principals and teachers. Some of the evaluation grantee analyses in the report are limited to 10 of the evaluation districts because 2 of the evaluation districts were not prepared to conduct random assignment of schools until the end of the school year in spring 2012. Those 2 districts were not administered principal and teacher surveys in 2012.

Data on district characteristics. To compare characteristics—such as students’ race and ethnicity, students’ eligibility for free and reduced-price lunch, average district enrollment, and geographic information—of all TIF 2010 districts with those of U.S. districts, the study team used information from the Common Core of Data (2009–2010).

Data on TIF implementation in all districts. To describe TIF program features and implementation experiences of TIF districts in general, the study team administered a survey to all TIF district administrators in December 2011.

Additional data on TIF implementation in 10 evaluation districts. The study team supplemented data obtained from the district surveys with information obtained through telephone interviews and technical assistance documents to describe in more detail TIF programs and implementation experiences in evaluation districts. The team conducted telephone interviews with staff in evaluation districts (such as the TIF program manager or director) in summer 2012. Technical assistance documents included needs assessments conducted in fall 2010 and spring 2011, and communication materials used by districts and grantees during the 2010–2011 planning year.

Data on teachers’ and principals’ attitudes and behaviors in 10 evaluation districts. The study team used teachers’ and principals’ survey responses to examine their understanding of the TIF program in their districts and to estimate the impact of pay-for-performance bonuses on their attitudes and behaviors. These surveys were administered to all principals in the evaluation schools and a sample of teachers in treatment and control schools in spring 2012. The teacher sample included all 1st- and 4th-grade teachers, and 7th-grade math, English language arts, and science teachers.

Methods

The study team’s analysis of the broad implementation of TIF for 2010 grantees relies on responses to the district survey. By calculating means or percentages, as appropriate, and giving equal weight to each district, we describe implementation of TIF in all districts and compare the experiences of evaluation districts with those of non-evaluation districts. To assess the alignment between educators’ understanding of the program and reports from evaluation districts, we compared mean responses for each of the three groups of survey respondents—districts, principals, and teachers.

To estimate the impact of the pay-for-performance component of TIF on educators’ attitudes and behaviors in evaluation districts, we compared survey responses of educators in treatment and control schools. We also conducted analyses separately by subgroups (such as how districts measured educators’ performance, the maximum value of their pay-for-performance bonuses, and teaching assignments) to assess how impacts on educators’ behaviors differed by program characteristics.

Study Sample and Characteristics of TIF Districts

Study sample. The final study sample for this report consisted of 153 TIF 2010 grantee districts, composed of 141 non-evaluation districts and 12 evaluation districts. For 10 evaluation districts, we also provide information about the experiences, behaviors, and attitudes of educators. The evaluation districts include 137 study schools in which all principals and a sample of 826 teachers were administered surveys.

Characteristics of TIF districts compared with all U.S. districts. The characteristics of the 2010 TIF districts are important for understanding the local contexts and types of districts interested in implementing the performance-based compensation system required by the TIF grant. Compared with all U.S. districts, TIF districts were significantly larger, were more likely to be located in urban

areas, had a higher proportion of disadvantaged and minority students, were more heavily located in the South, and were less likely to be in states with collective bargaining requirements.

Summary of Findings

The analyses presented in this report are based on information collected during the 2011–2012 school year, the first year of TIF implementation for most districts. By this time, districts had designed and communicated their performance-based compensation systems. Educators should have been provided information on the program’s components but they had not yet received (1) information on how they fared on their districts’ performance measures for the 2011–2012 school year or (2) any performance-based bonuses. Thus, the views and experiences of educators are based on only part of the process playing out, that is before receipt of any possible bonuses.

Key Findings About All 2010 TIF Districts and Their Programs

This section describes implementation findings from all 153 TIF districts. These results are based on the district survey administered in spring 2012.

Fewer than half of districts reported implementing all required components of the TIF program.⁴ Although 85 percent of TIF districts reported implementing at least three of the four required components for teachers, 46 percent reported implementing all four required program components for teachers. While most districts only had difficulty implementing one of the four required components, that component varied by district.

Most TIF districts generally met the grant requirements for measuring educator effectiveness. More than 80 percent of TIF districts reported using student achievement growth to evaluate teachers, 95 percent measured teacher effectiveness based on at least two formal classroom observations, 90 percent reported using student achievement growth to evaluate principals, and 75 percent reported using observations to evaluate principals. The approaches TIF districts used to measure student achievement growth varied. Most frequently, TIF districts reported measuring achievement growth for the entire school (76 percent) to evaluate teachers, followed by measuring individual teachers (69 percent) and subgroups of teachers (48 percent). Forty-two percent of TIF districts used growth measures at all three levels to evaluate teachers.

On average, expected pay-for-performance bonuses were 4 percent of the average U.S. educator’s salary, and the expected maximum pay-for-performance bonus was approximately double the average bonus. Districts also expected to provide some bonus to more than 90 percent of teachers and principals. The average TIF district expected to pay an average pay-for-performance bonus of \$2,462 to teachers in grades and subjects subject to annual accountability testing and a maximum bonus of \$5,355. Districts expected to award an average principal bonus of \$3,888 and a maximum bonus of \$6,282. The average TIF district expected that 93 percent of teachers in tested grades and subjects and 95 percent of principals would earn bonuses.

⁴ The TIF application notice also required grantees to collect and evaluate additional forms of evidence to measure educator effectiveness as part of the core elements to support program implementation. We do not include it here because it was not one of the four prioritized program components. However, in Chapter III we describe the types of additional evidence used by grantees.

Most TIF districts offered teachers additional pay opportunities, but fewer offered such opportunities to principals. Most TIF districts (87 percent) reported offering teachers additional pay opportunities, particularly for serving as mentor teachers (66 percent) or master or lead teachers (55 percent). Master and mentor teachers were offered the largest incentives—an average maximum of \$7,145 for master or lead teachers and \$3,735 for mentor teachers. Only 15 percent of TIF districts reported offering principals incentives to take on additional responsibilities.

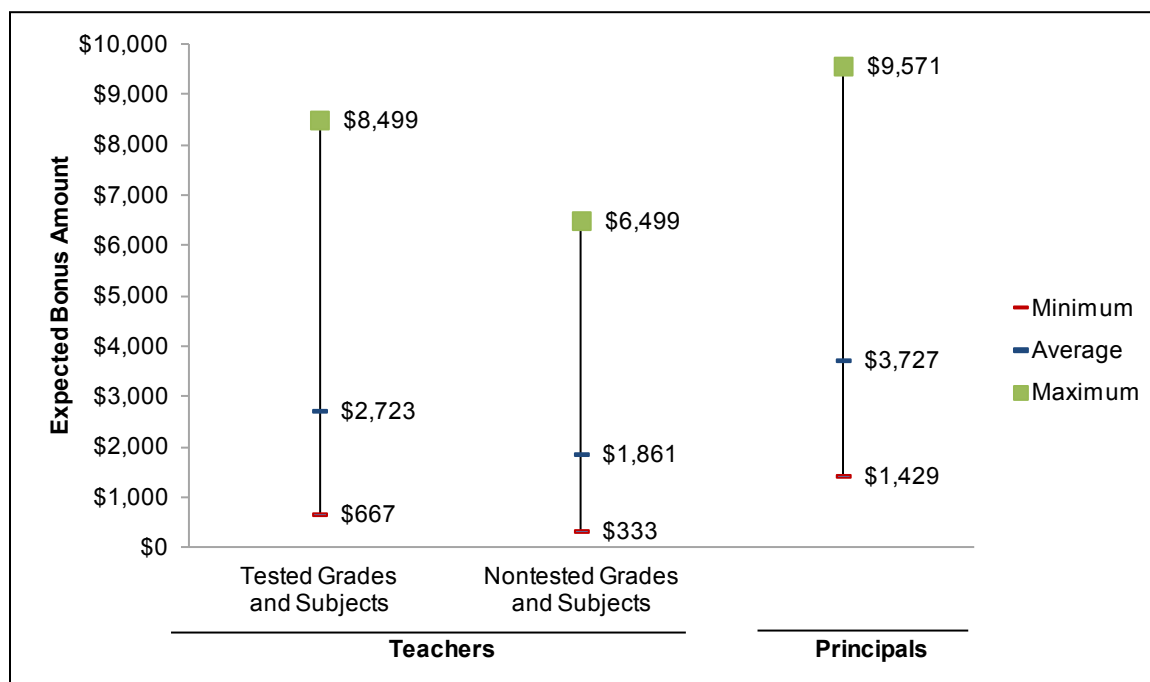
Key Implementation Findings for Evaluation Districts

This section describes implementation findings primarily for 10 of the 12 evaluation districts. The first two findings are based on all 12 evaluation districts; the remaining findings exclude the 2 districts that were not prepared for random assignment until the end of the 2011–2012 school year. In addition to the district survey, these results are based on data collected from technical assistance documents, interviews with district staff, and surveys of teachers and principals.

About three-quarters of the evaluation districts implemented all of the required components of TIF for teachers. All the evaluation districts reported using at least two formal classroom observations and student achievement growth to measure teachers' effectiveness, and offering pay-for-performance bonuses and additional pay opportunities. Fewer evaluation districts implemented the component that required observations of principals—one-quarter did not conduct observations of principals using trained observers.

Consistent with the TIF grant goals, evaluation districts expected pay-for-performance bonuses to be substantial and differentiated. However, the districts expected most educators would receive a bonus, suggesting that the award criteria were not consistent with TIF guidance for challenging pay-for-performance bonuses. Evaluation districts expected average pay-for-performance bonuses to be 4.8 percent of the average U.S. teacher's salary, very close to the 5 percent provided as an example in the TIF grant notice. However, average bonuses for principals were expected to be 4.0 percent of the average U.S. principal's salary, lower than the 5 percent example in the grant notice. Evaluation districts expected their bonuses to be differentiated, with the maximum bonuses offered for teachers and principals 3.1 and 2.6 times greater than the average, respectively. Evaluation districts expected that more than 75 percent of educators would receive some type of bonus, suggesting that the bonuses were not consistent with guidance in the TIF evaluation competition notice to offer payments only to those who perform significantly better than average. In Figure ES.2, we show the maximum, average, and minimum pay-for-performance bonuses that evaluation districts expected for teachers and principals.

Figure ES.2. Expected Pay-for-Performance Bonuses for Teachers and Principals in Evaluation Districts, Averages Across Districts

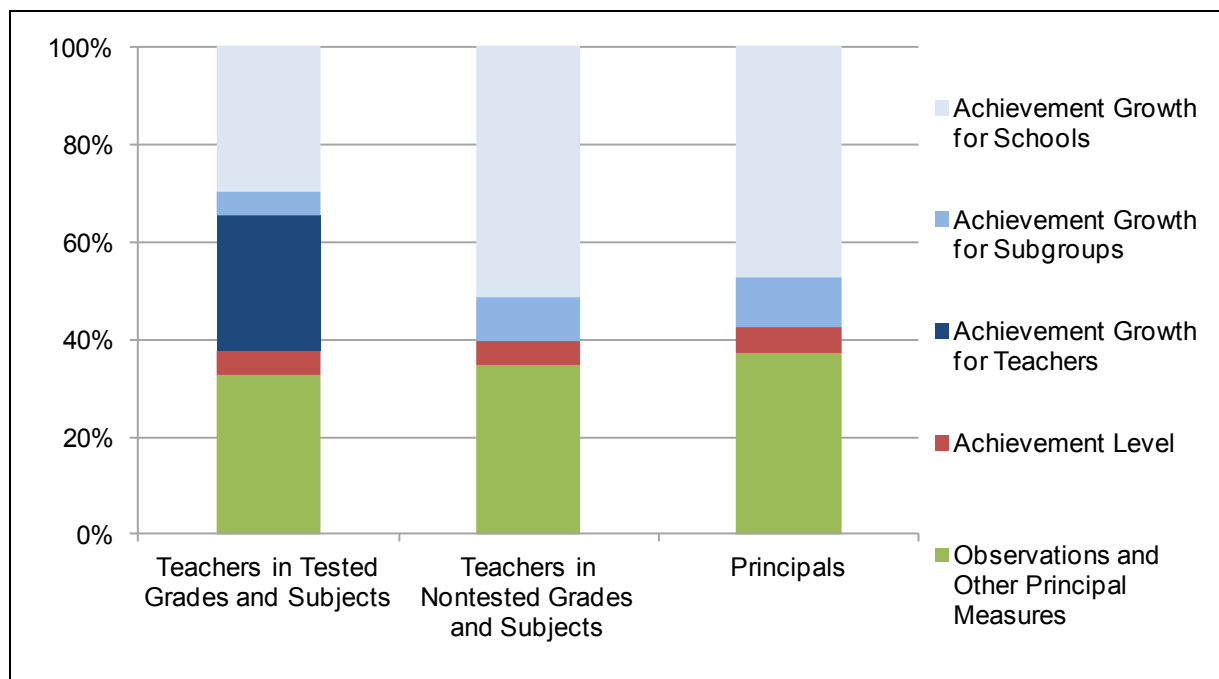


Source: District survey administered to evaluation districts.

Note: Based on survey questions about the expected distribution of TIF-funded pay-for-performance bonuses, given 10 categories of bonus amounts ranging from \$0 to \$15,000 or more. Although surveys were administered to all evaluation districts, only six of 12 were able to answer questions about the expected range of pay-for-performance bonuses for teachers and principals.

Evaluation districts offered separate bonuses for different types of performance measures. Teacher bonuses for performance based on student achievement growth were larger than bonuses for performance based on classroom observations. Eight of the 10 evaluation districts offered a separate bonus for each type of achievement growth measure (for example, one bonus for student achievement growth for the whole school and one for student achievement growth in a teacher's classroom) and a separate bonus based on classroom observations of teachers. The two remaining districts used a classroom observation measure to determine teachers' eligibility for a bonus based on achievement growth. On average, bonuses based on achievement growth comprised more than half of the expected total bonuses for teachers (62 percent) and principals (55 percent). In Figure ES.3, we show the relative weight of each type of performance measure for evaluation districts.

Figure ES.3. Relative Weight of Each Type of Performance Measure Used for Pay-for-Performance Bonuses in TIF Evaluation Districts



Source: Technical assistance documents.

Note: Ten evaluation districts. Because some evaluation districts combined a principal observation measure with other measures, such as surveys of teachers and parents, we combine these measures into one category for principals.

In evaluation districts, educators' reported awareness of evaluation measures often differed from districts' reports; principals' reports were more consistent with districts' reports. About two-thirds (68 percent) of teachers reported being evaluated on student achievement growth measures, 78 percent of teachers reported being evaluated through formal observations, and 89 percent of principals reported being evaluated on the basis of student achievement growth for their entire school (Table ES.1). In contrast, all of the evaluation districts reported using these measures to evaluate teachers and principals in TIF schools.

Teachers and principals in treatment schools reported lower rates of eligibility for pay-for-performance bonuses and lower expected pay-for-performance bonuses than districts reported. Figure ES.4 shows the maximum pay-for-performance bonuses expected by teachers and principals, and the maximum pay-for-performance bonuses districts reported they expected to award to teachers and principals. Although all teachers and principals in treatment schools were eligible for pay-for-performance bonuses, fewer than half (48 percent) of the teachers and 55 percent of principals in treatment schools thought they were eligible. On average, teachers in treatment schools perceived that the maximum pay-for-performance bonus was about \$2,800—less than a third of the maximum amount evaluation districts expected to offer teachers. Even among teachers in treatment schools who thought that they were eligible for pay-for-performance bonuses, the teachers believed that the maximum amount was about \$5,800. On average, principals in treatment schools thought that they could earn up to about \$4,700 in pay-for-performance bonuses—less than half the amount evaluation districts expected to award to treatment principals.

Table ES.1. Performance Measures Used to Evaluate Teachers and Principals, as Reported by Educators and District Representatives

	Percentage of Respondents Reporting the Measure Was Used		
	Teacher Report	Principal Report	District Report
Teacher Performance Measures			
Student achievement growth	68.0*+	56.3*	100.0
Classroom observations	78.1*+	97.5	100.0
Sample Size—Range^a	809–811	133–134	10
Principal Performance Measure			
Student achievement growth for the school	n.a.	88.7*	100.0
Sample Size	n.a.	127	10

Sources: Teacher, principal, and district surveys.

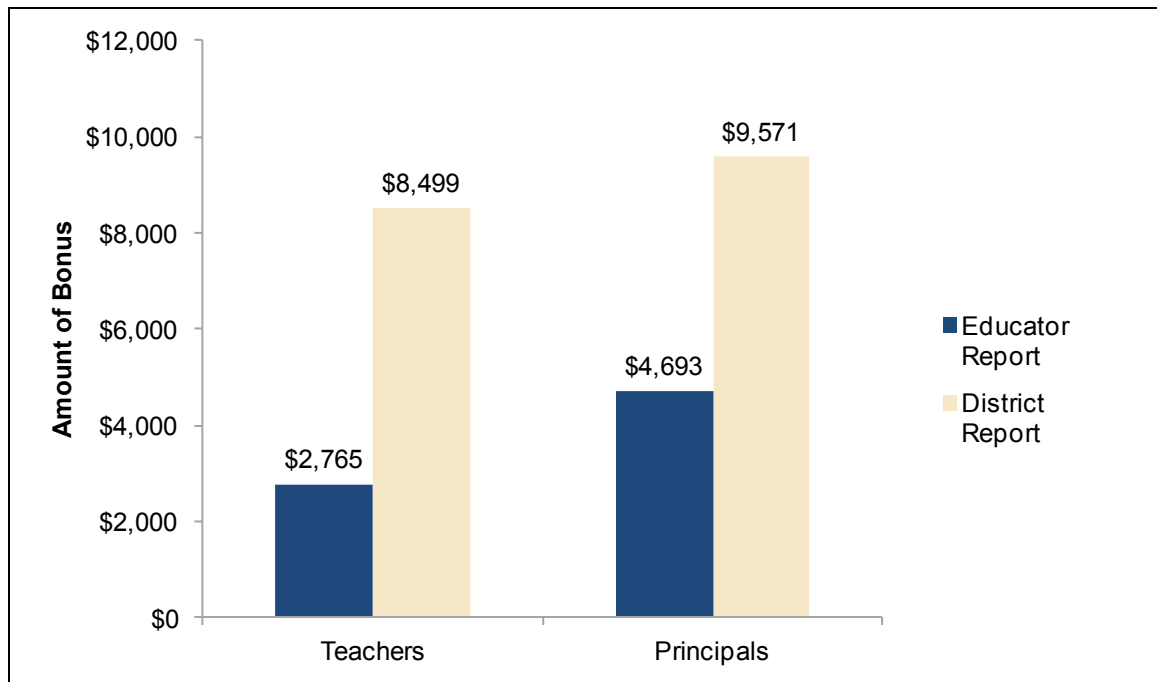
^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference from the district report is statistically significant at the 0.05 level, two-tailed test.

+Difference between teacher and principal reports is statistically significant at the 0.05 level, two-tailed test.

n.a. = not applicable.

Figure ES.4. Maximum Possible Size of Pay-for-Performance Bonuses for Treatment Teachers and Principals, as Reported by Respective Educators and Districts



Sources: Teacher, principal, and district surveys.

Note: Figures indicate respondents' average report of the maximum possible size of teachers' or principals' pay-for-performance bonuses. A total of 395 treatment teachers and 67 treatment principals responded to this survey question from 10 of the evaluation districts.

Many teachers were also not aware that they could earn additional pay for additional responsibilities. Only 61 percent of teachers reported that they or their colleagues in the same school were eligible to earn additional pay for extra responsibilities, even though all evaluation districts reported offering this type of additional pay.

Key Findings on Impacts of Pay-for-Performance Bonuses on Educators' Attitudes and Beliefs for Evaluation Districts

Finally, we summarize key findings on the impact of pay-for-performance on the attitudes and beliefs of teachers and principals in study schools within the 10 evaluation districts in which teacher and principal surveys were administered in 2012. We asked teachers and principals generally about their satisfaction with evaluation measures, professional opportunities, and school environment, as well as about their attitudes specifically toward TIF.

Most teachers and principals in treatment and control schools reported being satisfied with their professional opportunities, performance measures, and school environment. More than 65 percent of teachers reported being satisfied with how they were evaluated, and about 70 percent of teachers were satisfied with their jobs overall. About 85 percent of principals reported being satisfied with the feedback on their performance.

A lower percentage of teachers in treatment schools than in control schools reported that they were satisfied with performance measures, professional opportunities, school environment, and the TIF program, but a higher percentage were satisfied with their opportunities to earn extra pay. As shown in Table ES.2, teachers in treatment schools were less satisfied than teachers in control schools, on average, with the use of classroom observations as an evaluation measure, their professional opportunities, the quality of interaction with their colleagues, and school morale. Fewer teachers in treatment than control schools believed that TIF was fair and that it increased their job satisfaction. Teachers in treatment schools were also more likely to respond that the TIF program caused them to feel increased pressure to perform. The overall pattern of lower satisfaction among teachers in treatment schools compared with their control school counterparts had one exception: teachers in treatment schools were more satisfied with opportunities to earn extra pay.

A lower percentage of principals in treatment schools than in control schools reported that they were satisfied with school morale and with colleagues' contributions to student learning; yet principals' attitudes toward the TIF program were similar. Principals in treatment schools reported significantly lower satisfaction with school morale than principals in control schools (71 versus 88 percent) and were less likely to be satisfied with colleagues' contributions to student learning (94 versus 100 percent). However, there was no difference in principals' attitudes toward TIF, such as whether TIF contributed to greater teacher collaboration and whether the TIF program had been clearly communicated to them.

Teachers in treatment schools reported spending more time on instruction than teachers in control schools, but not more time overall on other activities during school hours. Teachers in treatment schools reported that they spent 48 minutes more on classroom instruction in the most recent full week of teaching than teachers in control schools. However, the difference in time spent on all activities—including supervising students, preparation time, and professional development—was not statistically significant.

Table ES.2. Teachers' Satisfaction with Performance Measures, Professional Opportunities, School Environment, and TIF Program

	Treatment	Control	Impact
Attitudes Toward Aspects of Teaching	Percentage Who Are Somewhat or Very Satisfied		
Classroom Observations as an Evaluation Measure	68.4	77.0	-8.6*
Opportunities for Professional Advancement	67.8	75.7	-7.8*
Quality of Interaction with Colleagues	73.6	80.6	-7.0*
School Morale	48.1	54.9	-6.8*
Opportunities to Earn Extra Pay	64.0	58.9	5.1*
Number of Teachers—Range^a	405–408	405–412	
Attitudes Toward TIF Program	Percentage Who Agree or Strongly Agree		
My Job Satisfaction Has Increased Due to the TIF Program	27.1	32.0	-4.9*
The TIF Program Is Fair	53.0	57.6	-4.6*
I Feel Increased Pressure to Perform Due to the TIF Program	62.9	54.1	8.7*
Number of Teachers—Range^a	399–403	394–403	

Source: Teacher survey.

^aSample 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.

Principals in treatment schools reported that TIF changed the way they recruited teachers to their schools, but not how they assigned staff within schools. Although principals in treatment and control schools tended to emphasize similar points to recruit teachers, more treatment than control school principals (1) used pay-for-performance bonuses to recruit teachers (26 versus 17 percent), and (2) used the TIF program as a recruiting incentive (46 versus 29 percent). For most measures, principals did not indicate they used different criteria to assign teachers to grades and subjects. Principals in treatment schools were significantly less likely (by about 10 percentage points) than those in control schools to report using a teacher's ability to raise test scores in their staffing decisions.

A small percentage of teachers or principals overall reported that TIF influenced their career choices, but teachers in treatment schools were more likely to report that TIF influenced their choice of where to teach. Fewer than 6 percent of teachers and 14 percent of principals reported that TIF affected their choice of schools. However, more treatment than control school teachers reported that TIF affected where or what they taught (5.5 versus 3.6 percent). Treatment school principals were even more likely than control school principals to report that they stayed at their current school because of TIF (10 versus 3 percent).

Looking Forward

This study was designed to provide implementation information for the 2010 TIF grantees. For the subset of grantees that received evaluation grants, the evaluation examines the impact of pay-for-performance bonuses as part of a comprehensive reform system within a large, multisite random assignment study design. Targeted technical assistance supported program implementation in the 12 evaluation districts to help ensure the proper implementation of their performance-based compensation systems.

Because educators' understanding of and responses to this policy may change over time, this study plans to follow the districts throughout the five-year grants. In addition to examining any changes in the findings presented here, future reports will examine the impact of the pay-for-performance component on student achievement and educator mobility after one or more years of TIF implementation.

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 students' 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 first of four planned reports from a multiyear study focusing on the TIF grants awarded in 2010.⁵ It examines grantees' implementation experiences and intermediate educator outcomes near the end of the first year of program implementation, before the first pay-for-performance payouts to teachers and principals. Future reports will address the impacts of such payouts on student achievement, educator mobility, and changes in educators' job satisfaction and attitudes toward their TIF programs.

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 affect the retention and recruitment of educators and, ultimately, student achievement.

Previous Research on Pay-for-Performance Programs for Educators

Current 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.⁶

Experimental Evidence

Most experimental studies found no statistically significant impact of pay-for-performance on student achievement. Three studies (Marsh et al. 2011; Fryer 2011; Goodman and Turner 2010) examined the impact of New York City's School-Wide Performance Bonus Program, whereby teachers could earn a bonus of up to \$3,000 per full-time union member at the school. The studies found no overall impact of the program on student achievement, high school graduation rates, or teacher retention or absences, and in a few cases, researchers found small negative effects on math or reading achievement in certain years. A multiyear study using both a random assignment and

⁵ TIF grants are often 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.

⁶ For a more detailed discussion of the pay-for-performance literature, see Glazerman et al. (2011).

matched comparison group design found that the Chicago Teacher Advancement Program (TAP)—a comprehensive teacher pay reform model similar to the national TAP—did not raise student math or reading scores but did increase teacher retention in some schools (Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012).

Springer et al. (2010) examined the impact of the Project on Incentives in Teaching (POINT) program in Nashville, Tennessee on student math achievement in grades 5 through 8. POINT offered substantial pay-for-performance bonuses (\$5,000 to \$15,000) to middle school math teachers. The authors found no overall impact of pay-for-performance bonuses on student math achievement, although in the program’s second and third years, there were positive impacts on 5th grade students’ achievement. Another study, Springer et al. (2012), examined the impact of team-level teacher awards in middle schools in Round Rock, Texas, and found no statistically significant impact on student achievement or teacher-reported practices or attitudes.

In a study by Fryer et al. (2012), teachers from nine schools in Chicago Heights, Illinois, were randomly assigned to one of three groups:

- A traditional pay-for-performance program in which teachers received bonuses at the end of the year based on their students’ achievement gains (the “Gain” group)
- A group in which teachers received a lump-sum payment at the beginning of the year that they would return if their students did not meet performance targets (the “Loss” group)
- A control group that was not eligible for a performance-based bonus

The authors found statistically significant positive gains in students’ math achievement for the Loss group and no significant impact on students’ math achievement for the Gain group.

Nonexperimental Evidence

A recent study examined financial incentives for teachers rated highly effective by the District of Columbia’s new teacher evaluation system (Dee and Wyckoff 2013). Highly effective teachers earned a one-time bonus of up to \$25,000, and those rated highly effective for two consecutive years could earn a permanent salary increase from \$6,000 per year to more than \$20,000 per year. In comparing teachers just above and below the cutoff for earning a highly effective rating, the study found that the incentives did not increase teacher retention but did improve teacher performance (as measured by the district’s evaluation system).

Several recent studies have examined the association between pay-for-performance bonuses—offered through TIF grants as well as other large or national programs—and student achievement growth, using study designs that did not rely on random assignment of teachers or schools. Springer et al. (2008) compared gains in student test scores between schools that implemented TAP and those that did not. The authors found that TAP had a positive impact on student test-score gains at the elementary level, but a negative impact for grades 6 to 10. Studies on pay-for-performance incentive programs in Texas found no systematic association between the programs and student achievement or teacher turnover (Springer et al. 2009a, 2009b). A study of the performance-based compensation system implemented in Charlotte-Mecklenburg, North Carolina schools found that students’ reading and math achievement grew more quickly in TIF schools than in comparison schools over the five-year grant period (Slotnik et al. 2013). A study of Houston’s ASPIRE program (Shifrer et al. 2013) found that teachers’ receipt of a bonus was associated with gains in student

achievement and higher teacher retention and attendance. Because the ASPIRE program included nearly all Houston schools, the study could not examine whether the program itself was associated with better outcomes. Finally, Bayonas (2010) examined the impact of Mission Possible, a comprehensive performance-based compensation program implemented in 28 schools in Guilford, North Carolina. (Eight of these implementations were funded by a previous TIF grant.) The study found no differences between treatment and comparison schools in students' reading or math achievement.

Although evidence is growing, there are still few high quality studies of comprehensive, well-implemented pay-for-performance programs. Thus, 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 include the following:

- **Study design limitations.** The studies that do not rely on random assignment 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 of 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. Although many have argued that pay-for-performance bonuses may improve the teaching workforce by increasing the chances of recruiting and retaining effective teachers, studies have rarely been designed to examine the impact of pay-for-performance bonuses on one or both of these outcomes (Fryer et al. 2012; Shifrer et al. 2013; Springer et al. 2010; Springer et al. 2012; Glazerman et al. 2009; Glazerman and Seifullah 2010, 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, bonus amounts varied little with performance, and teachers received similar bonuses regardless of their measured effectiveness (Marsh et al. 2011; Fryer 2011; Goodman and Turner 2010; 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 2010; Shifrer et al. 2013). In addition, communication about the program was in some cases very limited (Springer 2010).
- **Other design features of pay-for-performance programs that may influence educator and student outcomes.** Many of the existing studies examined programs offering other features that may strengthen or weaken the influence of pay-for-performance on educators. For example, 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 programs examined the impact of pay-for-performance within the context of these more comprehensive reforms, which included features such as support for effective teachers to take on leadership roles or to participate in professional development activities (Glazerman et al. 2009; Glazerman and Seifullah 2010, 2012; Bayonas 2010; Slotnik et al. 2013; Springer et al. 2008); others did not (Fryer et al. 2012; Springer et al. 2010; Marsh et al. 2011; Fryer 2011; Goodman and Turner 2010). Similarly, the criteria for earning pay-for-performance bonuses may affect the impact 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 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 both 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. This evaluation was designed as a large, multisite random assignment study of the impact on educators and students of pay-for-performance as part of a comprehensive reform system. In addition, program implementation was supported by targeted technical assistance to help ensure programs were well designed. 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.⁷ Among the 62 TIF grantees in 2010, more than two-thirds were states or school districts (69 percent), 16 percent were non-profits, 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 partially supported 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 program 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 program 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. It was also necessary that the evaluation give significant weight to student achievement growth—defined as the change in student achievement for an

⁷ 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>.

individual 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 to 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 program 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 who 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 were also required to have the proper supports to implement and maintain the performance-based compensation system. The following five core elements were required:

1. **The involvement and support of teachers, principals, unions (if applicable), and other personnel needed to carry out the TIF grant.** Grantees were required to involve educators in the design of the performance-based compensation system and to demonstrate educator support before implementation. Examples of support could be letters from superintendents, results of votes in favor of the program by school staff, or signed agreements with unions.
2. **A rigorous, transparent, and fair evaluation system for teachers and principals.** The evaluation system had to include student achievement growth measures and at least

two observations per year by trained observers; it also had to differentiate among educators. Grantees were asked to demonstrate the internal capacity to implement these measures including (1) calculating measures of student achievement growth rates annually based on state accountability exams, and (2) training staff to reliably administer a rigorous, transparent tool for observations of teachers and principals. The core elements also required that grantees collect additional forms of evidence—beyond achievement growth and observations—to incorporate into the evaluations of teachers and principals.

3. **A plan to effectively communicate the components of its performance-based compensation system.** An essential part of a performance-based system is communication to teachers, administrators, other school personnel, and the community at large. Grantees had to consider how best to communicate with stakeholders through, for example, printed materials, face-to-face meetings, or webinars. Grantees also needed to provide suitable staff support to present the information to stakeholders as well as enough time for educators to digest the information and ask questions.
4. **A plan for ensuring educators understood the measures of educator effectiveness.** Through staffing and technology, grantees had to develop appropriate infrastructure to provide educators with information about their performance on the effectiveness measures of the performance-based compensation system. Grantees also had to develop professional development programs to help educators improve along the effectiveness measures.
5. **A data management system that could link student achievement data to educator payroll and human service systems.** For example, data systems needed the capacity to store all data compiled on effectiveness measures and then link them to a payroll system so that bonuses could be awarded.

The required components of the performance-based compensation system are comprehensive and designed to work together, so it was necessary that grantees have the core elements in place before implementing their compensation systems. Grantees without 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. 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 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. The criteria for earning a bonus based on the effectiveness measures could also vary, such as 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 in 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. State education agencies, local education agencies, charter management organizations, and nonprofit organizations could apply to one or the other. Evaluation grant applicants had to agree to participate in a more in-depth evaluation of their TIF grants. Specifically, there were three additional evaluation grant requirements:

- **Random assignment.** One of the most important requirements was that applicants for an evaluation grant 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). Districts were allowed to offer educators in control schools an across-the-board bonus of no more than 1 percent of salary. This bonus was not tied to effectiveness measures but was intended to solidify participation in the study.
- **Minimum number of schools.** Grantees were required to include at least eight elementary or middle schools in the evaluation.
- **Data collection.** Grantees were obligated to cooperate with all data collection activities for the evaluation.

Another key difference between the main and evaluation grant requirements is that applicants for the evaluation grants were given more specific guidance about the structure of their pay-for-performance bonus. They received examples of a 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 for evaluation grants had discretion over the proposed structure of the pay-for-performance bonus, these examples provided additional guidance to applicants and may have influenced the design of their performance-based compensation systems.

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. In return, ED provided the grantees \$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. Differences in the requirements of the non-evaluation and evaluation grants, as well as in the additional guidance provided, may have resulted in systematic differences between the structure of the performance-based compensation system for evaluation and non-evaluation grantees. We examine these potential differences in Chapter IV.

Designing and implementing a performance-based compensation system is challenging; therefore, ED took several steps to help ensure that this evaluation studied the policy as it was envisioned, rather than as a program that was experiencing startup problems or partial implementation in its first year. These steps included:

- Specifying program-design requirements
- Requiring applicants to justify their design choices
- Requiring grantees to demonstrate they had met the required five core elements prior to implementing their performance-based compensation system
- Providing technical assistance to grantees

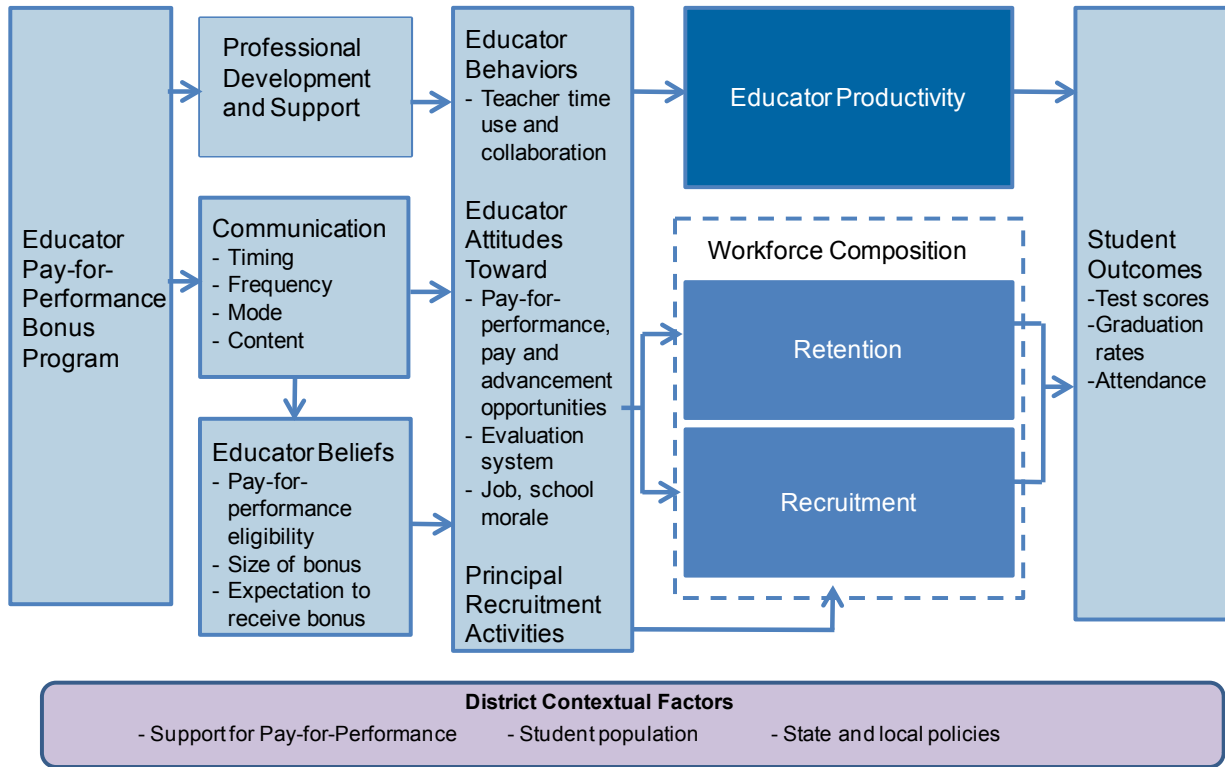
ED monitored both evaluation and non-evaluation grantees to ensure implementation was consistent with grant requirements. Although ED ensured all grantees received technical assistance, it used two providers—one for the non-evaluation grantees and one for the evaluation grantees. Resources for the evaluation grantee technical assistance helped ensure that the evaluation grantees received intensive and targeted assistance. The evaluation 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 affect 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 action of how pay-for-performance might lead to improved student outcomes. We developed a logic model to show the pathways by which pay-for-performance 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, as hypothesized by the model that informed the data collected as part of the evaluation.

Districts adopt a pay-for-performance program that rewards educators based on their measured effectiveness. The ability to earn a pay-for-performance bonus, and the fact that the criteria to earn a bonus depends on student achievement gains, can potentially impact teachers' attitudes toward their school choice, alter their teaching practices, and increase their productivity. For example, pay-for-performance programs may serve as incentive for effective teachers to remain in a school that provides bonuses and may also 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 who are rewarded for student achievement gains on standardized tests may allocate more time to instructional practices intended to improve test scores.

Figure I.1. Logic Model



However, whether and how pay-for-performance programs 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 a pay-for-performance program only if they understand how they are being evaluated and how they can change their teaching practices to improve their performance. They must also 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 informs the emphasis on these activities required by the grant.

Educators’ understanding of the pay-for-performance program will depend on the districts’ specific communication activities, the timing of communication, and the 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 fail to attend informational meetings or read the materials offered. Furthermore, educators must be made aware of the program when there is still sufficient time either to affect their school choice (for example, request a school transfer) or to alter their teaching practices.

The ability of a pay-for-performance program to impact educator behaviors and attitudes also depends on the district context, such as educators’ support for the program 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

inciting effective educators to leave the school.⁸ 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.

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 five research questions:

1. What are the characteristics of all TIF grantee 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. What is the impact of pay-for-performance bonuses on students' achievement on state assessments of math and reading?
4. How do pay-for-performance bonuses affect educator mobility, including whether mobility differs by educator effectiveness?
5. What performance-based compensation system features are associated with student achievement or educator mobility?

This study includes information on implementation of TIF for all 2010 grantees (question 1) and more in-depth implementation and impact information from a subset of 12 districts selected through the evaluation competition (questions 2 through 5). In this first report, the study team focuses on early implementation of the TIF grants (questions 1 and 2), specifically, the features of districts' performance-based compensation systems, the structure of the pay-for-performance bonuses, and educators' understanding of their districts' programs. In addition, the study team examines the impact of pay-for-performance bonuses on intermediate outcomes related to educators' attitudes, productivity, recruitment, and retention near the end of the first year of implementation.

⁸ 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 positively impact productivity. See Kamenica (2012) for a review.

The analyses presented in this report are based on information collected from TIF districts and from principals and teachers in evaluation districts during the 2011–2012 school year, the first year of implementation for most districts. At that point in time, districts had designed and communicated their performance-based compensation systems. Educators should have had an opportunity to learn about the program’s components but had not yet received information on how they performed on the measures of effectiveness for the 2011–2012 school year, nor had they received any bonus based on their performance. Thus, the views and experiences of educators are based on only part of the process playing out.

Road Map for the Remainder of the Report

In the remainder of this report, we describe in detail the study’s design and findings. In Chapter II, we describe the study sample, the design of the experimental evaluation, the data used for this report, and the analytic approaches. In Chapter III, we describe the characteristics of all 2010 TIF districts, their TIF programs, and their experiences implementing TIF. In Chapter IV, we provide more detailed information about implementation experiences in the subset of 12 TIF evaluation districts, and in Chapter V we examine how eligibility for pay-for-performance bonuses impacted teachers’ and principals’ attitudes and behaviors.

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, and present an overview of the analytic approaches.

Study Sample

As we explained in Chapter I, the TIF 2010 grant notice included both a main competition and an evaluation competition. In this report, we refer to the subset of TIF grantees and districts that were awarded a grant to participate in the random assignment evaluation as “evaluation grantees” and the subset that did not apply to the evaluation competition as “non-evaluation grantees.”⁹ In 2010, the Department of Education (ED) awarded 62 TIF grants, which included 183 districts. Of these 62 grants, 11 were evaluation grants, which included 15 districts.¹⁰

In Table II.1, we list the original number of grants awarded in 2010, the number of districts included in those TIF grants, and the number that continued to participate in TIF during the 2011–2012 school year.¹¹ We also list the number of grantees, districts, and schools in the analyses for this report (“the analysis sample”). The analysis sample excludes 3 evaluation and 9 non-evaluation districts that did not implement TIF in the 2011–2012 school year, as well as 18 non-evaluation districts that did not respond to the district survey. In addition, it excludes educators in 39 study schools in two evaluation districts where the study team did not administer teacher and principal surveys, as described below.

The final study sample for this report consisted of 153 TIF districts, which included 141 non-evaluation districts and 12 evaluation districts. In this report, we describe key program characteristics and implementation experiences of these 153 districts, but we focus on the implementation experiences of the subset of 12 evaluation districts. We also provide information about the experiences, behaviors, and attitudes of the educators in 137 study schools within the evaluation districts, including all principals and a subset of 826 teachers.

Characteristics of 2010 TIF Districts Compared with All U.S. Districts

The characteristics of TIF districts are important for understanding the local context and the types of districts interested in implementing the performance-based compensation system required by the TIF grant. For example, the TIF notice’s requirement to focus on higher poverty schools or the resources required to develop and implement a comprehensive performance-based compensation system may be more feasible for certain types of districts. Here, we describe TIF districts compared with the average U.S. district using data available from the Common Core of Data.

⁹ All TIF districts are included in the study to provide information on the implementation of TIF across all districts. The evaluation districts are participating in the in-depth evaluation, which is examining the details of program implementation as well as the impact of pay-for-performance bonuses on educator and student outcomes.

¹⁰ One evaluation grantee is an association of charter schools in Michigan. For simplicity, we treated this consortium as one school district.

¹¹ Some grantees were unable to implement their programs and withdrew from the grant; we describe these circumstances in Chapter III.

Table II.1. TIF Grants Awarded, Grantees Implementing TIF in 2011–2012 School Year

	Awarded TIF Grant in 2010	Implemented TIF in 2011– 2012 School Year	Analysis Sample
Grantees and Districts			
Grantees			
All grantees	62	56	56
Evaluation grantees	11	9	9
Districts			
All TIF districts	183	171	153 ^a
Evaluation districts	15	12	12
Schools			
Study Schools Within Evaluation Districts	250	176	137 ^b

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

^aDistrict analyses included information on the 153 TIF districts that implemented TIF in 2011–2012 and responded to the district survey.

^bWe administered the teacher and principal surveys to educators at 137 study schools (located in 10 of the 12 evaluation districts). Because two evaluation districts had not finalized their TIF programs by spring 2012, random assignment of their 39 study schools occurred in summer 2012. Therefore, we did not administer the educator surveys to the teachers and principals in these 39 schools in spring 2012.

The 2010 TIF districts differed from average U.S. districts in several ways. As we show in Figure II.1, TIF districts were larger—approximately 21,000 students enrolled, on average, compared with an average of 3,000 across all U.S. districts. TIF districts were also more likely to be located in urban areas and had a higher proportion of disadvantaged and minority students, with a higher free or reduced-price lunch (FRL) eligibility rate than all U.S. districts (64 versus 47 percent). TIF districts were also more likely to be located in the South and less likely to be in states with collective bargaining requirements (Appendix A, Table A.1). The higher percentage of disadvantaged students in TIF districts is consistent with the requirement that districts implement TIF in high-need schools with at least 50 percent of students eligible for FRL.

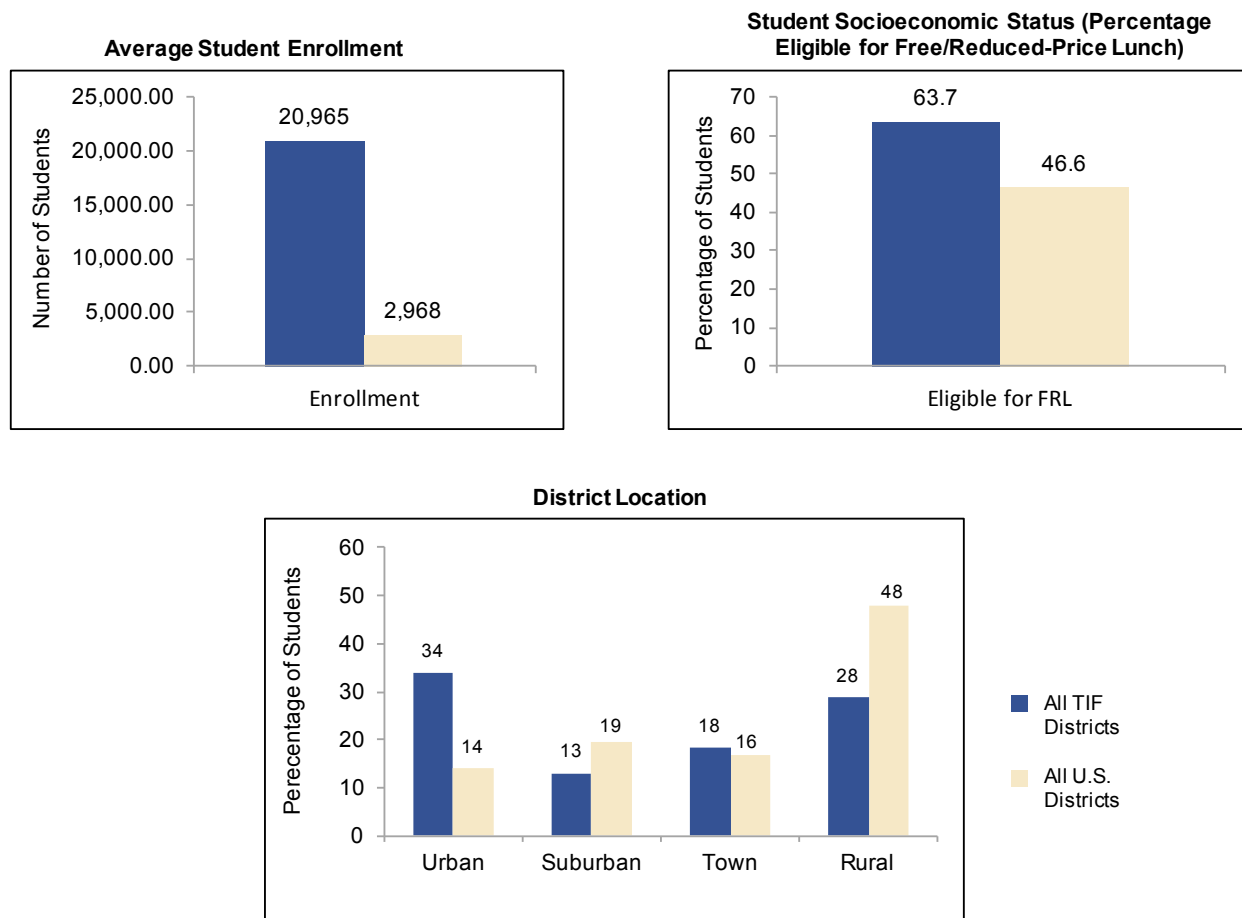
The districts that obtained a TIF grant, the schools included in those grants, and the subset of districts and schools that participated in the in-depth evaluation were not nationally representative of all U.S. districts, which has implications for the interpretation of the study findings. The set of all 171 TIF districts consists of districts that (1) demonstrated the need and desire to adopt a performance-based compensation system, as defined by the grant notice, and (2) undertook a successful grant-writing effort. Moreover, among all TIF grantees, the subset of 12 evaluation districts is neither nationally representative nor statistically representative of all TIF districts. Rather, it is composed of districts that agreed to meet additional requirements to receive an evaluation grant. Nevertheless, estimates based on evaluation districts can provide strong causal evidence of the effects of performance-based compensation system reform in a more diverse set of districts than in previous studies.

Characteristics of TIF Evaluation and Non-Evaluation Districts

Although ED used the same criteria to award evaluation and non-evaluation TIF grants, evaluation grantees and districts may differ from other TIF districts in important ways because of 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.

Figure II.1. Characteristics of TIF Districts Compared with All U.S. Districts



Source: Common Core of Data for 2009–2010 school year.

Note: All differences between TIF districts and U.S. districts are statistically significant at the 0.05 level.

There were no statistically significant differences in the characteristics of evaluation and non-evaluation districts. Given the relatively small sample size of 12 evaluation districts, only large differences are likely to be statistically significant, so we note differences that were larger than 10 percentage points or 10,000 students. As we show in Table II.2, evaluation districts were larger, on average, than non-evaluation districts. The evaluation districts were also more likely to be located in urban areas (67 versus 31 percent), in the West (42 versus 18 percent), and in states with collective bargaining (58 versus 34 percent), and less likely to be in the South (25 versus 45 percent) or Midwest (17 versus 29 percent). Although evaluation and non-evaluation districts were very similar in terms of the percentage of Hispanic students and students’ socioeconomic status, evaluation districts had a higher percentage than non-evaluation districts of black students (37 versus 25 percent) and a lower percentage of white students (53 versus 40 percent).

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	40.2	52.9
Black, non-Hispanic	37.0	24.9
Student Socioeconomic Status		
Eligible for free/reduced-price lunch	62.9	63.8
Size		
Number of students	35,037	19,676
District Location		
Urban	66.7	31.2
Suburban	16.7	12.8
Town	8.3	19.1
Rural	8.3	30.5
Geographic Region		
Northeast	16.7	8.5
Midwest	16.7	29.1
South	25.0	44.7
West	41.7	17.7
Collective Bargaining ^a		
In state with collective bargaining	58.3	34.0
In state without collective bargaining	41.7	66.0
Sample Sizes		
Districts	12	130
Grantees	9	44
States	8	23

Source: Common Core of Data for 2009–2010 school year.

Notes: Table is based on 142 of the 153 TIF districts that were included in the analyses, with 130 non-evaluation districts and 12 evaluation districts. Eleven non-evaluation districts were not included in the 2009–2010 district-level Common Core Data.

^aCollective bargaining is a state-level indicator from the National Right to Work Legal Defense Foundation (<http://www.nrtw.org/rtnws.htm>).

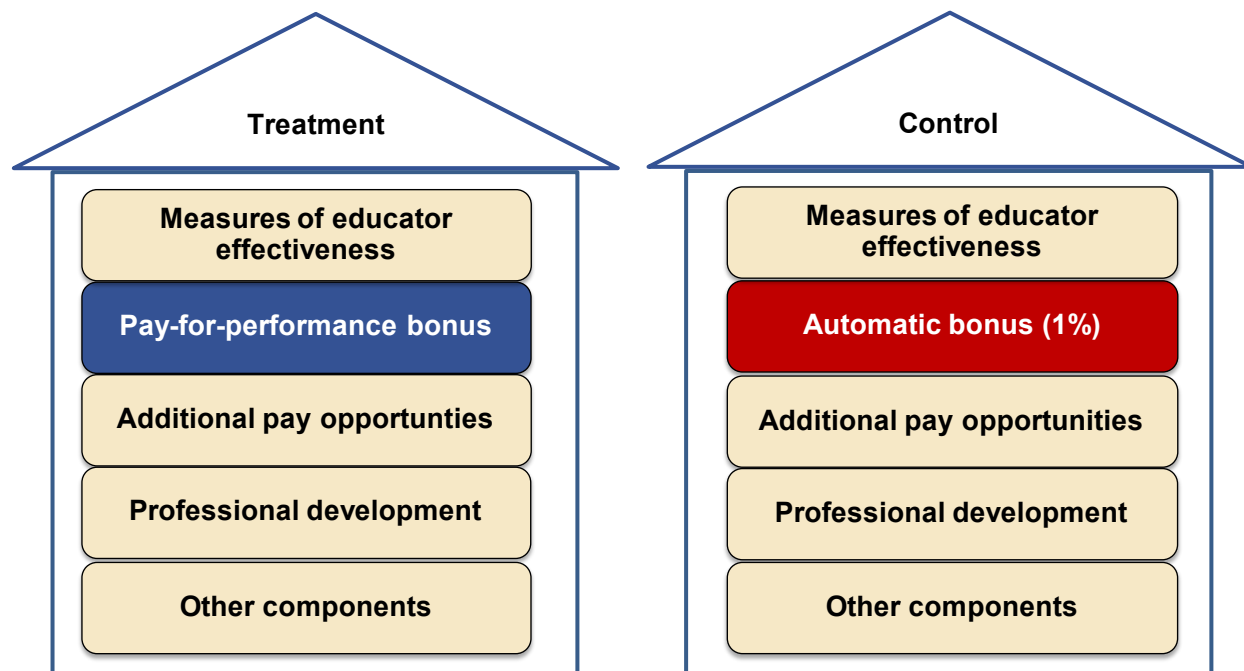
None of the differences between evaluation and non-evaluation TIF districts is significant at the 0.05 level, two-tailed test.

Experimental Design to Estimate the Impact of Pay-For-Performance

To ensure that inferences about the effect of pay-for-performance are based solely on the offer of pay-for-performance and not on other characteristics of districts, schools, or educators, we randomly assigned schools within a district to treatment and 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 TIF grant notice required the 1 percent bonus in control schools. The 1 percent bonus ensured that all educators in evaluation schools received some benefit from participating in the study, either a pay-for-performance bonus or the automatic bonus. Therefore, the impact of pay-for-performance estimated in this study is

based on two potential effects (i) bonuses in treatment school were differentiated based on educator performance, and (ii) bonuses in treatment schools were a little larger on average, than in control schools.

Figure II.2. Random Assignment Design



Prior to random assignment, schools were matched based on characteristics measured before the district's implementation of TIF—primarily prior 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). We describe random assignment procedures in more detail in Appendix A.

This random assignment process created two groups that, on average, should have included students and schools with similar characteristics and differed only in the opportunity for educators to receive pay-for-performance bonuses. In Appendix A, Table A.2, we show that treatment and control schools' baseline characteristics were similar in terms of school type, enrollment, and location, and student achievement, race and ethnicity, and socioeconomic status.

Several factors influenced the timing of random assignment. Because we wanted to examine how pay-for-performance influenced teachers' decisions to remain in or move from their current schools, we wanted to randomly assign schools early enough to allow teachers the opportunity to change schools. We did not, however, want to randomly assign schools so early that the study would lose schools because of unanticipated closures or changes in leadership. We also took into account district requests. For example, one district felt strongly that it needed to inform teachers about their school's assignment early enough to allow time for them to request a transfer. Three districts were randomly assigned by March 2011, seven were randomly assigned in May and June 2011, and the remaining two were randomly assigned in June and July 2012.

As we discuss in the next section, the analyses in this report mainly focus on the evaluation districts. We use detailed information from all 12 evaluation districts to describe their TIF programs and experiences. However, because 2 of the 12 evaluation districts were not prepared for random

assignment of their schools until summer 2012, we did not administer surveys to the principals and teachers in those districts in spring 2012. Therefore, the 137 study schools and the responses from the educators in the schools that are part of this report's analyses come from 10 of the 12 evaluation districts.

Data Sources

The analyses in this report are based on data from six main sources (see Table II.3):

1. The Common Core of Data (CCD)
2. A survey administered to all TIF districts
3. Interviews with TIF staff in evaluation districts
4. Principal surveys in evaluation districts
5. Teacher surveys in evaluation districts
6. Documentation from teams that provided technical assistance to the evaluation districts

Table II.3. Data Sources for First TIF Report

Data Source	Data Obtained	Timing of Data Collection	Sample	Mode
CCD	Composition of student characteristics in districts	2009–2010	All U.S. districts	n.a.
District Survey	TIF program features, implementation experiences	12/2011–6/2012	All TIF districts	Hard copy
District Interviews	Detailed information on TIF implementation and program features	6/2012–7/2012	TIF evaluation districts	Phone
Principal Survey	TIF program features, attitudes toward TIF program and job, hiring practices	3/2012–6/2012	Principals in schools in 10 evaluation districts	Web and hard copy
Teacher Survey	TIF program features, attitudes toward TIF program and job, time use	3/2012–6/2012	1st, 4th, 7th grade (math, ELA, and science) teachers in schools in 10 evaluation districts	Web and hard copy
TA Documents	Detailed information on implementation and program features	Fall 2010–spring 2012	12 evaluation districts	n.a.

Note: n.a. = not applicable.

Common Core of Data. We obtained information from the CCD on the characteristics of TIF districts and all U.S. districts. We compared TIF districts (overall and by evaluation status) with all U.S. districts on such characteristics as students' race and ethnicity, FRL eligibility, average district enrollment, and geographic information.

District survey. The district survey included questions about the district's experience implementing its TIF program, specifically the required components of its performance-based compensation system. We addressed these surveys to the individual identified as overseeing or directing the district's TIF program. We also compared the experience of evaluation districts with non-evaluation districts to examine the degree of similarity between their programs and experiences. This comparison reveals whether the more detailed implementation findings obtained from the evaluation districts may be relevant to the broader set of all TIF districts.

Within evaluation districts, we also compared district staff members' responses about components of their TIF programs with educators' responses to similar questions. This comparison examines whether educators' understanding of their TIF program aligns with that of the district staff members. Because clear communication is critical for a pay-for-performance program to influence educator practices and student achievement, this analysis provides an initial indication of some of the conditions necessary for the program to have an effect.

We administered the district survey to all districts (14 evaluation and 168 non-evaluation districts) that were awarded a 2010 TIF grant or were included as part of a state or other entity's 2010 TIF grant.¹² As we show in Appendix B, Table B.1, 91 percent of districts (151 non-evaluation and all 14 evaluation districts) responded to the district survey.¹³ In Appendix B, Table B.2, we compare characteristics of district respondents with nonrespondents. The groups are very similar on key characteristics, such as the school's student racial composition, student socioeconomic status, size, and location.

Although the overall district response rate was high, response rates to specific questions varied considerably. In particular, few districts responded to certain questions that related to the pay-for-performance bonuses. For example, fewer than 60 percent of eligible district respondents answered a question regarding the expected size and distribution of teacher bonuses. Respondents may have had difficulty with this question because the district survey was administered before districts awarded their first-round bonuses. In addition, the structure of some proposed pay-for-performance programs may have made it difficult to answer the closed-ended questions on the survey, leading some respondents to skip them.¹⁴ In contrast, almost all district respondents (151) could answer questions about principal performance measures.

Follow-up interview with evaluation districts. The interviews with TIF evaluation district administrators provided more in-depth information than that collected from the survey, and allowed

¹² We administered 182, rather than 183 district surveys, because we knew at the time we fielded the survey that one of the evaluation districts had dropped out of TIF. Therefore, we surveyed 14 of the original 15 evaluation districts.

¹³ The final analysis sample excludes 3 evaluation districts (including the one that was not surveyed) and 9 non-evaluation districts that did not implement TIF in the 2011–2012 school year, as a result, the final study sample consisted of 141 non-evaluation and 12 evaluation districts

¹⁴ For example, if a district offered a separate pay-for-performance bonus for each measure of educator performance, they may have had difficulty estimating the distribution of the bonuses without information on the expected distribution of teacher performance on each measure.

us to probe for more information on specific features of districts' TIF programs. For example, for a few districts, we clarified the criteria to obtain a bonus and the maximum expected bonuses. We also gathered more contextual background information related to districts' TIF implementation, such as the extent to which educators in the evaluation districts already faced strong performance-based incentives from other policies. Finally, we solicited more information on implementation challenges, strategies, and communication activities. We conducted follow-up phone interviews with all 12 evaluation districts that were still participating in TIF in spring 2012.

Principal and teacher surveys. Both the principal and teacher surveys asked respondents about their district's TIF program and their attitudes toward TIF and their job. For example, the surveys included questions about how educators were being evaluated, their eligibility for either a pay-for-performance or automatic 1 percent bonus, whether the program included incentives for educators to take on additional responsibilities in their schools, and professional development activities. Educators were asked about their satisfaction with the use of classroom observations, pay-for-performance bonuses, school morale, and collaboration with colleagues. They were also asked whether and how TIF may have influenced their decision to stay at or leave their current school.

Certain topics were specific to either the principal or teacher survey. For example, principals were asked about their hiring practices and approach to assigning teachers to grades and subjects. Teachers were asked about mentoring activities and how they allocated their time throughout the school day.

We used educator survey responses for three main purposes:

- To describe educators' understanding of their TIF program
- To assess the degree to which principals' and teachers' understanding and experiences were consistent with district responses
- To compare the experiences, attitudes, and classroom and school practices of educators in treatment and control schools

The only difference in the TIF program between treatment and control schools was the offer of pay-for-performance bonuses. If educators' understanding and experiences under the TIF program are similar for both groups, then we can attribute differences in their behaviors and attitudes to the offer of pay-for-performance bonuses.

We administered surveys to all principals and to a sample of more than 1,000 teachers in the 137 study schools. 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. These groups represent elementary and middle school grades and subjects both with and without annual accountability testing. In Appendix A, we explain in detail how we determined the teacher sample. Appendix B, Table B.3, shows the composition by grade and subject of the 826 teachers who met the selection criteria. As the table indicates, the teacher sample included similar numbers of 1st grade, 4th grade, and 7th grade teachers. Approximately 98 percent of principals (99 percent in the treatment schools and 97 percent in the control schools) responded to the principal survey, and 92 percent of teachers (91 percent in the treatment schools and 93 percent in the control schools)

responded to the teacher survey. There were no statistically significant differences between respondents and nonrespondents (see Appendix B, Tables B.7 and B.8).¹⁵

Technical assistance documents. The technical assistance team documented various aspects of the evaluation districts' programs and implementation activities and experiences.¹⁶ It conducted needs assessments in fall 2010 and spring 2011 with staff from each evaluation district or grantee. The assessments evaluated the following areas:

- Program design and planned implementation
- Progress in implementing the five core elements required by ED
- 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.

Overview of Analytic Approach

In this section, we discuss the analytic approaches used in Chapters III, IV, and V. In Chapter III, we focus on the implementation experiences of all TIF districts; in Chapter IV, on the implementation experiences of TIF programs in the evaluation districts; and in Chapter V, on the impact of pay-for-performance on teachers and principals in study schools. In Appendix C, we provide more technical details on the analytic methods, the primary analysis, and the sensitivity analyses performed.

Implementation of TIF in All Districts

To describe TIF implementation across districts, in Chapter III, we draw 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 reflect the experiences of the average district that implemented a TIF program.

Implementation of TIF in Evaluation Districts

In Chapter IV, we describe the implementation of TIF in the subset of districts that participated in the evaluation. In addition to the district survey, we used information collected from only the evaluation districts. Sources included technical assistance documents, district telephone

¹⁵ In Appendix B, Tables B.5 and B.6, we show response rates by district for teachers and principals, respectively. In Appendix B, Tables B.7 and B.8, we compare school characteristics of teacher and principal survey respondents with their respective full-survey samples. Although a few school characteristics showed significant differences between the teacher respondents and the full sample of teachers, the school characteristics were similar in magnitude. None of the school characteristics showed a significant difference between principal respondents and the full sample of principals.

¹⁶ The technical assistance team consisted of Mathematica staff and a consultant from Vanderbilt University who provided support to the evaluation districts to implement their TIF programs consistent with the specification of the TIF evaluation grant notice.

interviews, and teacher and principal surveys. We calculated means (or percentages, as appropriate) within the various groups of interest—for example, all non-evaluation and evaluation districts. We weighted each district equally, and looked at differences in implementation between non-evaluation and evaluation districts.¹⁷

As discussed in Chapter I, pay-for-performance programs can motivate educators to change their behaviors, or attract or retain effective educators, only if they are aware of their eligibility to receive a bonus. The success of the incentive program can also depend on its structure (such as the size of potential bonuses and the perceived likelihood of receiving one). Therefore, a key aspect of the implementation analysis was to assess whether educators' understanding of the program aligned with district reports. To determine this information, we compared mean responses for each of the three groups of survey respondents: (1) districts, (2) principals, and (3) teachers.¹⁸

Impact of Pay-for-Performance on Teachers and Principals in Evaluation Districts

The final aspect of TIF implementation that we examine in this report is the impact of the pay-for-performance component of TIF on educators' attitudes and behaviors. These results reflect interim outcomes that could affect student achievement, as described in the logic model discussed in Chapter I. They include overall job satisfaction, satisfaction with the TIF program, the desire to remain in or leave their current school, and use of time in the classroom and throughout the day.

We estimated this impact by analyzing the difference in these outcomes between treatment and control schools within the same districts. Because the study used random assignment, any differences in educators' attitudes or behaviors can be attributed to pay-for-performance and not some other characteristic of the districts, schools, or educators. To estimate these impacts, we used a linear regression model that adjusted for the random assignment design—in particular, the assignment of groups of educators within schools rather than individual educators, as well as the pairing of these clusters before random assignment. First, we estimated the district-specific impact, in which we weighted schools equally. Next, we calculated the average impact in the full study sample by taking a weighted average of the district-specific impacts. The district-specific impacts were weighted by the number of schools in the evaluation.¹⁹ We tested the robustness of the impact findings to a variety of alternative methods including the choice of weights, method for calculating standard errors, and specification of regression equations with binary outcomes. In Appendix C, we describe these sensitivity analyses in more detail. The results of the sensitivity tests are discussed along with the corresponding primary findings for each outcome in Chapter V.

We also conducted analyses separately by subgroups to determine whether impacts on educators' behaviors differed by these characteristics. We created subgroups based on characteristics of the TIF programs and characteristics of the teachers. As part of the analysis of the program subgroup, we grouped districts into three categories based on how heavily a district weighted student achievement growth measured at the classroom level in its teacher-evaluation measures (see

¹⁷ We used two-sided *t*-tests (or chi-squared test for categorical outcomes) to assess statistical significance.

¹⁸ Because the number of principal and teacher respondents differed by district, we compared responses across teachers, principals, and district staff within districts and then averaged estimated differences across districts. We used two-sided *t*-tests to determine whether differences between any two types of respondents were statistically significant, and *F*-tests to jointly test differences between the three types of survey respondents.

¹⁹ The mean outcome for the treatment group was calculated as the unadjusted mean outcome of the control group plus the regression adjusted difference in outcomes between the two groups.

Chapter IV for more detail). We also grouped districts by the distribution of expected payouts to determine whether the impacts differed based on the size of districts' maximum bonuses. As part of the teacher subgroup analysis, we looked at differences in impacts by teaching assignment (that is, teachers in grades and subjects with annual accountability testing compared with teachers with no annual testing), and years of teaching experience.

THIS PAGE IS INTENTIONALLY BLANK

III. TIF DISTRICTS AND THEIR PROGRAMS

In this chapter, we present a broad picture of TIF programs in 2011–2012 by examining the four required program components of the TIF grant, discussed in Chapter I. We first describe TIF districts' reports about each required component of TIF. Next, we examine how many TIF districts implemented all of the required components. We conclude the chapter with details about how districts involved educators in their programs' designs.

The findings presented in this chapter are based on reports from 153 districts that (1) were included in TIF 2010 grants, (2) implemented a TIF program in the 2011–2012 school year, and (3) completed a district survey. District staff completed the surveys between December 2011 and March 2012. The timing was approximately halfway through the 2011–2012 school year, which for almost all districts was the first year of program implementation.

Key Findings About TIF Districts and Their Programs

- **Fewer than half of districts reported implementing all required components of the TIF program.** Although 85 percent of TIF districts reported implementing at least three of the four required components for teachers, slightly fewer than half (46 percent) reported implementing all four.
- **Most TIF districts (80 percent) met the grant requirements to use student achievement growth and multiple observations to measure educator effectiveness.**
- **Consistent with the TIF grant goals, grantees expected pay-for-performance bonuses to be somewhat substantial and differentiated. However, the districts expected most educators would receive a bonus, suggesting that the award criteria were not consistent with TIF guidance for challenging pay-for-performance bonuses.** TIF districts expected to award an average pay-for-performance bonus of about 4 percent of the average U.S. educators' salary. The maximum bonus expected by TIF districts was twice as large as the average bonus for teachers and 50 percent larger than the average bonus for principals. Districts also expected to award a pay-for-performance bonus to more than 90 percent of eligible teachers and principals.

Requirement 1 -- Measures of Educator Effectiveness

The TIF notice required that districts measure educator effectiveness using student achievement growth and at least two observations of educators per year by trained observers. Within those requirements, districts had substantial flexibility in choosing how to (1) assess student achievement growth, (2) evaluate classroom or professional practices, and (3) use the performance measures to determine effectiveness. In this section, we describe the measures TIF districts used to evaluate educators, focusing on whether districts used student achievement growth and observations.

More than 80 percent of TIF districts reported using student achievement growth to evaluate teachers. The approaches they used, however, varied. Some used the achievement growth of students in a teacher's own classroom, resulting in each teacher receiving a different score. Some used achievement growth of an entire grade level to determine the score for all teachers at that grade level. Others used achievement growth for the entire school based on all teachers in tested grades and subjects (Table III.1). Many districts used a combination of these three approaches. Most frequently, TIF districts reported measuring achievement growth for the entire school (76 percent),

followed by individual teachers (69 percent), and subgroups of teachers (48 percent). Forty-two percent of TIF districts used all three types of growth measures to evaluate teachers. Finally, 45 percent of TIF districts also reported using a measure of the level of student achievement, such as average test scores or proficiency rates at one point in time (although no districts reported using only achievement levels).

Table III.1. Percentage of Districts Using Measures of Student Achievement to Evaluate Teachers

	All TIF Districts
Student Achievement Measures	
Any achievement measure	83.3
Achievement growth	83.3
By schools	76.0
By student subgroup ^a	48.3
By teachers' classrooms	69.3
Achievement level	45.3
Number of Districts—Range^b	149–150

Source: District survey.

^aStudent subgroups can be defined by grade, teams, subject areas, and demographic characteristics.

^bSample sizes are presented as a range, based on the data available for each row in the table.

Almost all TIF districts reported conducting at least two formal classroom observations.

Ninety-five percent of TIF districts measured teacher effectiveness based on at least two formal observations, as required by the grant. TIF districts planned to conduct, on average, four or five formal observations of teachers, lasting about 43 minutes each (Table III.2). Because this number of observations would require a substantial amount of staff time, TIF districts typically relied on principals and other staff to conduct observations. Principals or administrators conducted observations in 95 percent of TIF districts. In two-thirds of TIF districts, at least one other type of staff member also conducted classroom observations (not shown): teacher leaders—such as mentors or master teachers—or peer observers conducted observations in 54 percent of districts, and content specialists did so in 19 percent of districts. Eight percent used district administrative staff; 4 percent hired observers (Table III.2).

Table III.2. Classroom Observations to Evaluate Teachers in TIF Districts

	All TIF Districts
Classroom Observations	
Percentage of districts conducting classroom observations	98.0
Percentage of districts conducting at least two formal observations	95.2
Average number of observations per school year	4.6
Average length of observations in minutes	42.8
Percentage of districts in which observations are conducted by:	
Principals or other administrators at the teacher's school	95.3
Teacher leaders or peer observers	54.2
Content specialists	18.7
District administrative staff	8.0
Externally hired observers	4.0
Number of Districts—Range^a	146–153

Source: District survey.

^aSample sizes are presented as a range, based on the data available for each row in the table.

The TIF application notice also required grantees to collect and evaluate additional forms of evidence to measure educator effectiveness as part of the core elements to support program implementation. Districts reported that principals' professional judgment was the measure most frequently used (69 percent) for evaluation (Appendix D, Table D.1). Other measures used to evaluate teachers included teacher participation in school activities (41 percent) and teachers' attendance records (34 percent).

Ninety percent of TIF districts reported using student achievement growth to evaluate principals, and 75 percent reported using observations to evaluate principals.²⁰ Although all of the districts that used student achievement growth used it for the entire school (Table III.3), they sometimes combined it with other measures. For example, 52 percent of all districts combined school achievement growth and achievement growth among student subgroups, such as disadvantaged students. Districts could also use measures other than student achievement and observations to evaluate principals. Among the other measures, the most commonly used was teacher assessments of principal performance (48 percent of districts), and 15 percent of districts incorporated parent input (Appendix D, Table D.1).

Table III.3. District Report About Principal Evaluation Measures (Percentages)

Performance Measure of Principals	All TIF Districts
Student Achievement Measures	
Any achievement measure	91.5
Achievement growth	90.2
By schools	90.2
By student subgroups ^a	52.3
Achievement level	65.4
Observations	
Observations by trained observers	74.5
Number of Districts—Range^b	151–153

Source: District survey.

^aStudent subgroups can be defined by grade, teams, subject areas, and demographic characteristics.

^bSample sizes are presented as a range based on the data available for each row in the table.

Requirement 2 -- Pay-for-Performance Bonus

The TIF notice included a requirement that districts offer pay-for-performance bonuses based purely on a teacher's or principal's performance. These requirements indicated that criteria for earning a bonus and the structure of the bonus awards should be substantial, differentiated, and challenging. However, the notice did not provide formal definitions for these criteria. Examples provided in only the guidance to evaluation grant applicants suggested the following descriptions:

- **Substantial**—an average bonus equivalent to at least 5 percent of the average teacher salary
- **Differentiated**—a maximum bonus worth three times the average bonus
- **Challenging**—a bonus awarded only to educators who perform significantly better than average

²⁰ We did not collect in the district survey information about the length or frequency of the observations of principals. We provide more information about principal observations in the 12 evaluation districts in Chapter IV.

In this section, we report on the percentage of districts that offered pay-for-performance bonuses and the expected pay-for-performance bonus amounts. We describe these amounts as “expected,” because districts were still in the first year of implementation of their programs and had not yet awarded bonuses. Therefore, districts reported on the pay-for-performance bonuses they anticipated paying out based on educator performance. We use the examples of “substantial,” “differentiated,” and “challenging” that were given to evaluation grantees as benchmarks in examining the expected pay-for-performance bonus amounts offered by TIF districts. In a future report, we will describe the actual amounts awarded in evaluation districts.

As we discuss in Chapter II, some districts had difficulty answering questions about the expected distribution of the pay-for-performance bonus amounts earned by educators.²¹ Part of the reason for this difficulty may be that the district survey was administered before districts had evaluated teachers and awarded pay-for-performance bonuses for the 2011–2012 school year. This timing meant that districts had to estimate the maximum, average, and minimum bonus they expected to pay. Thus, the information presented here is based on fewer districts than the information presented on performance measures and other aspects of the design of districts’ TIF programs. For example, 87 of the 153 districts responded to a question about expected pay-for-performance bonuses for teachers, and 99 of the 153 districts responded regarding expected pay-for-performance bonuses for principals.²² Thus, the findings on the expected payouts may not generalize to all 2010 TIF districts.

Almost all of the TIF districts expected to offer pay-for-performance bonuses to teachers. Ninety-four percent of TIF districts reported that teachers in their districts were eligible for bonuses or awards based on their performance in the 2011–2012 school year (not shown). The same percentage of TIF districts reported that principals were eligible for bonuses or awards based on their performance in that school year.

The average TIF district expected to award an average pay-for-performance bonus that was about 4 percent of the average U.S. teacher salary. In Figure III.1, we show the maximum, average, and minimum expected bonus amounts reported by TIF districts for teachers of grades and subjects with and without annual accountability testing (referred to as tested grades and subjects), and for principals. The analysis included the following highlights:

- The average TIF district expected to pay an average pay-for-performance bonus of \$2,462 for teachers in tested grades and subjects (or 4 percent of the average U.S. teacher salary in 2011–2012 of \$57,000),²³ and a maximum pay-for-performance bonus of \$5,355.

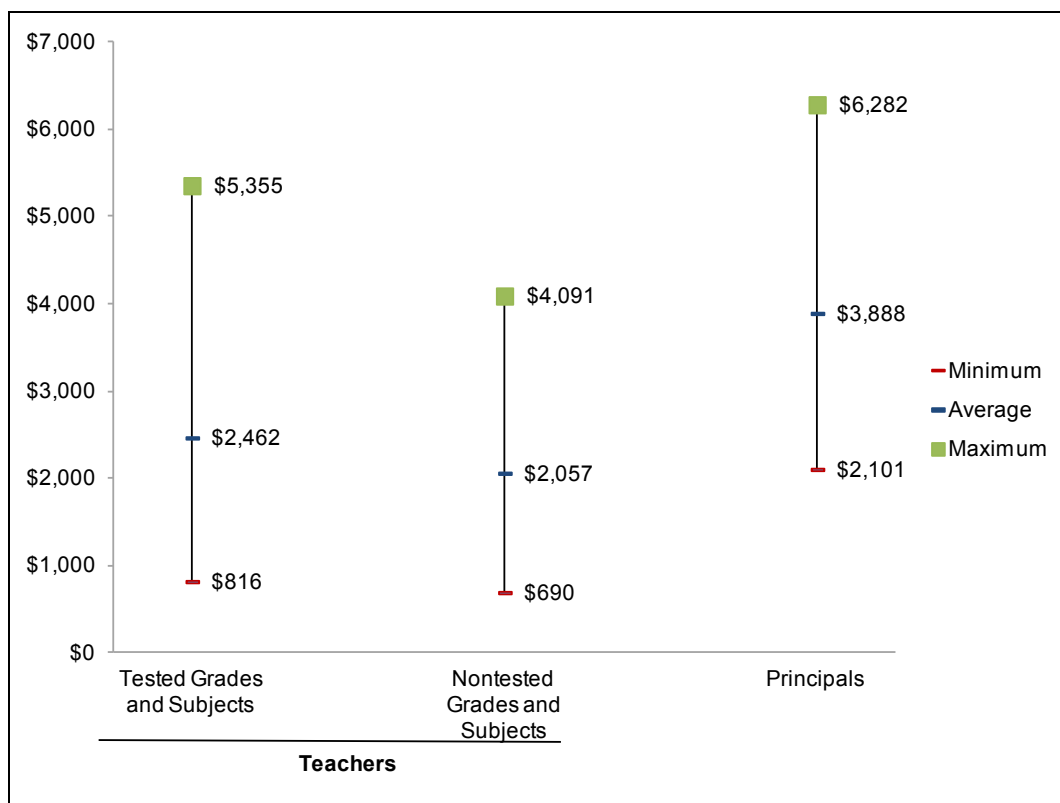
²¹ Predicting the distribution of pay-for-performance bonuses requires information on how educator performance would be distributed on the evaluation measures. As a result, districts using new evaluation measures may have had limited data to predict the distribution of educator performance on the evaluation measures.

²² In Appendix B, Table B.9, we show that districts that did and did not respond to questions about expected pay-for-performance bonuses were similar with regard to the student achievement measures used to evaluate teachers and prior experience with pay-for-performance bonuses. However, districts that did not respond were significantly less likely (45 percent of respondents versus 27 percent of nonrespondents) to implement the Teacher Advancement Program (TAP). They were significantly more likely (10 percent of respondents versus 22 percent of nonrespondents) to report that they revised their TIF program to better align with data management systems.

²³ These percentages are based on the average U.S. teacher’s salary, as reported in the U.S. Department of Education’s School and Staffing Survey (http://nces.ed.gov/programs/digest/d12/tables/dt12_083.asp).

- For teachers of nontested grades and subjects, TIF districts expected to pay lower pay-for-performance bonus amounts than for teachers in tested grades and subjects: an expected average bonus of \$2,057 and an expected maximum bonus of \$4,091.
- On average, TIF districts expected to pay a minimum bonus of \$816 for teachers in tested grades and subjects and \$690 for teachers of nontested grades and subjects.
- TIF districts expected to pay larger pay-for-performance bonuses to principals than to teachers. On average, districts expected an average principal bonus of \$3,888 (about 4 percent of the average U.S. principal salary of \$93,000)²⁴, and a maximum bonus of \$6,282.
- On average, TIF districts expected to offer principals a minimum bonus of \$2,101.

Figure III.1. Average, Minimum, and Maximum Expected Pay-for-Performance Bonuses for Teachers and Principals



Source: District survey.

Notes: The figure is based on answers to a question about the expected distribution of pay-for-performance bonuses, given 10 categories of bonus amounts that range from \$0 to \$15,000 or more (for example, the percentage of teachers expected to earn a bonus between \$1,000 and \$1,999). Eighty-seven of the 153 TIF districts responded to the question for teachers; 99 of the 153 districts responded to the question for principals. The maximum bonus by district was calculated as the top range of the largest category with a positive percentage of teachers or principals expected to receive a bonus in that range. The minimum bonus by district was calculated as the bottom range of the lowest category with a positive percentage of teachers or principals expected to receive a bonus in that range. The average bonus by district was calculated as the average of the midpoint dollar amount of each category, weighted by the percentage of teachers or principals expected to received a bonus in that range.

²⁴ These percentages are based on the average U.S. principal’s salary, as reported in the U.S. Department of Education’s School and Staffing Survey (http://nces.ed.gov/programs/digest/d12/tables/dt12_083.asp).

The average TIF district expected teachers to earn a maximum pay-for-performance bonus that was approximately two times the average bonus. The grant notice indicated that pay-for-performance bonuses should be differentiated; instead of most teachers receiving the same amount, bonuses should vary. One way to measure the amount of variation in pay-for-performance bonuses is to compare the maximum amount teachers were expected to earn with the average amount. As shown in Figure III.1, district staff expected the maximum bonus for teachers to be twice the amount of the average bonus. For principals, however, the expected maximum pay-for-performance bonus was only 1.5 times the average bonus.

Another way to measure variation in pay-for-performance bonuses is to examine the percentage of teachers and principals expected to earn different bonus amounts. In Figures III.2 and III.3, we display the percentages of teachers and principals that districts expected to receive varying amounts of pay-for-performance bonuses. These figures show that 79 percent of teachers and 60 percent of principals were expected to earn pay-for-performance bonuses of \$1 to \$3,999.

Although the expected distribution of bonus amounts for teachers in tested subjects varied little across TIF districts, we examined how the expected distribution of bonuses for these teachers varied within districts. The details of this analysis are in Appendix D, Figure D.1, but we highlight some of the key points here. Districts varied in the expected distribution of pay-for-performance bonuses. For example, some districts expected a fairly small range of pay-for-performance bonuses: a minimum of \$1,000, an average of \$1,500, and a maximum of \$2,000, for instance. Other districts expected more differentiated bonuses: a minimum of \$0, an average of about \$3,000, and a maximum of \$15,000. Minimum expected performance bonuses ranged from \$0 to \$6,000.

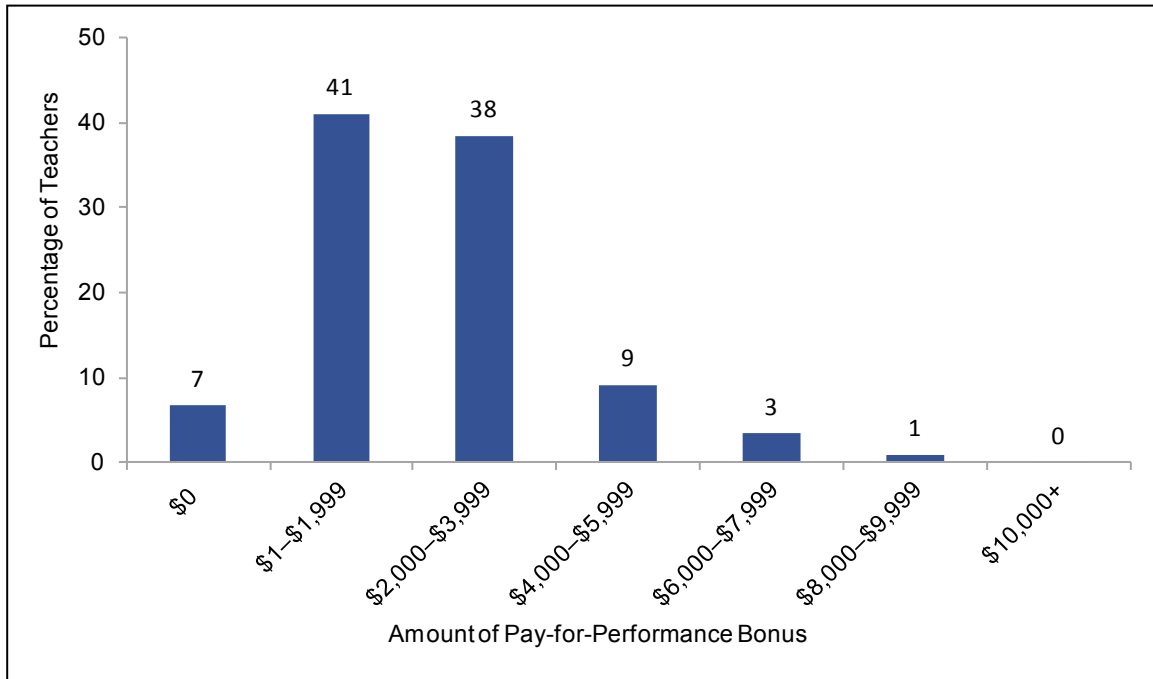
The average TIF district expected that 93 percent of teachers in tested grades and subjects, and 95 percent of principals would earn a pay-for-performance bonus. To examine whether TIF districts offered challenging bonuses, in Figures III.2 and III.3, we present the percentages of teachers in tested grades and subjects and principals that TIF districts expected would obtain a pay-for-performance bonus. Those figures—93 percent of teachers and 95 percent of principals—suggest that districts did not expect it would be difficult for educators to obtain pay-for-performance bonuses.

Requirement 3 -- Additional Pay Opportunities

Consistent with the goal of improving the teaching workforce in high-need schools, the TIF notice required that programs provide financial incentives for educators to take on additional roles and responsibilities. The TIF notice also encouraged applicants to offer additional pay for educators to teach in high-need subject areas or to work in hard-to-staff schools by recruiting effective educators into these positions or retaining the educators already filling these positions. In this section, we examine the expected pay amounts for taking on these additional responsibilities. We also contrast these amounts with those offered as pay-for-performance to provide context for the various incentives available to educators.

Eighty-seven percent of TIF districts reported offering teachers additional pay opportunities. Most commonly, TIF districts reported offering additional pay for teachers to serve as mentors (66 percent) or master or lead teachers (55 percent) (Table III.4). Less-commonly offered opportunities included additional pay for teachers to serve as department chairs, as leadership team members, on a schoolwide committee, or as a lead curriculum specialist. Prior to TIF, most TIF districts (53 percent) provided teachers additional pay for taking on enhanced responsibilities in their schools, such as assuming the role of a mentor, master, or lead teacher (Appendix D, Table D.2).

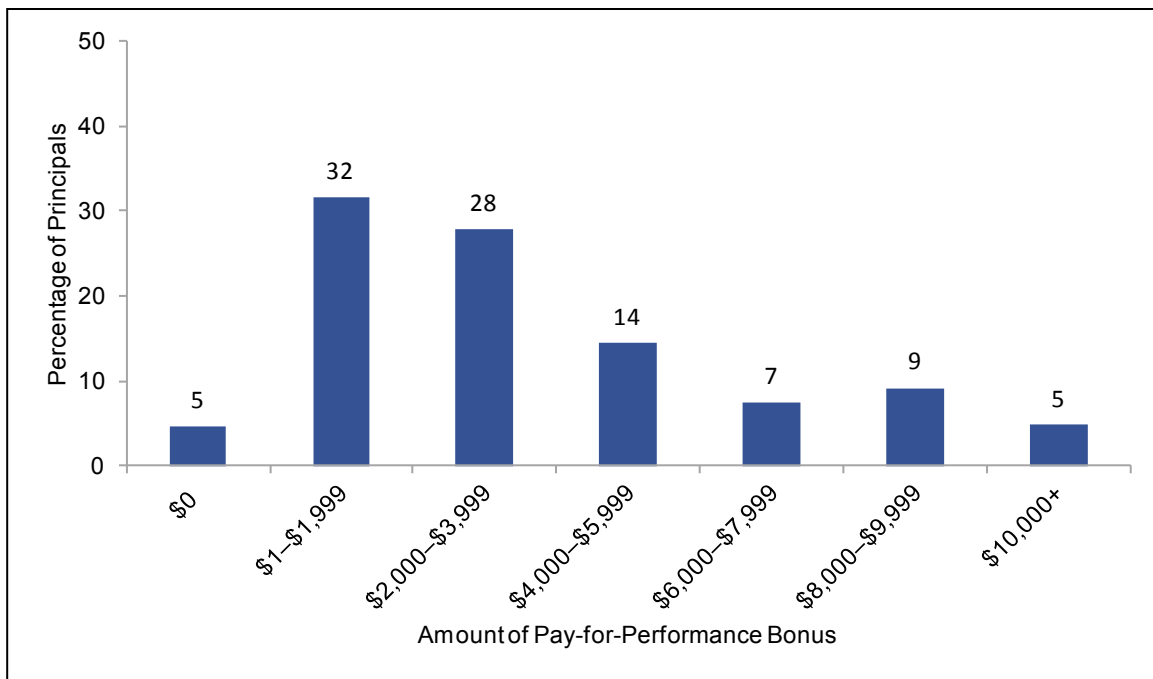
Figure III.2. Expected Distribution of Teacher Pay-for-Performance Bonuses in Tested Grades and Subjects



Source: District survey.

Note: 87 districts responded to this survey question.

Figure III.3. Expected Distribution of Principal Pay-for-Performance Bonuses



Source: District survey.

Note: 99 districts responded to this survey question.

Teachers who agreed to become master and mentor teachers were offered the largest incentives. The average maximum amount offered for master or lead teachers was \$7,145; the average maximum for mentor teachers was \$3,735 (Table III.4).²⁵ TIF districts that offered additional pay for other types of roles or responsibilities, such as serving as a leadership team member or lead curriculum specialist, reported amounts of \$1,107 to \$2,320.

A minority of TIF districts elected to offer additional pay opportunities to principals. For example, 15 percent offered principals incentives to take on additional responsibilities; 11 percent reported offering principals additional pay—an expected average of \$5,212—for working in a hard-to-staff school (Table III.4).

Table III.4. Additional Pay Opportunities for Teachers and Principals

	Percentage of TIF Districts That Offered Additional Pay	Maximum Amount of Additional Pay in Districts Offering It
Teachers		
Assuming Additional Responsibilities	86.6	n.a.
Roles and Responsibilities		
Mentor teacher	66.2	\$3,735
Master or lead teacher	55.1	\$7,145
Department chair or head	22.3	\$1,416
Lead curriculum specialist	8.9	\$2,320
Schoolwide committee or task force member	16.9	\$1,256
Leadership team member	23.4	\$1,107
Additional Factors		
Teaching in hard-to-staff school	17.4	\$3,602
Teaching in high-need subject area	23.6	\$3,455
Attending professional development activities or enrolling in graduate-level courses	27.8	\$780
Number of Districts—Range^a	144–149	10–88
Principals		
Assuming Additional Responsibilities in School or District	15.4	\$2,206
Additional Factors		
Working in hard-to-staff school	11.2	\$5,212
Attending professional development activities or enrolling in graduate-level courses	16.1	\$838
Number of Districts—Range^a	143	13–18

Source: District survey.

Note: Table reports on activities funded by TIF.

^aSample sizes are presented as a range based on the data available for each row in the table.

n.a. = not applicable.

²⁵ The district survey did not explicitly define the terms “mentor teacher” or “master teacher.” The TIF grant notice said that additional responsibilities and leadership roles were “additional duties teachers may voluntarily accept, such as serving as master or mentor teachers, who are chosen through a performance-based selection process (including through assessment of their teaching effectiveness and the ability to work effectively with other adults and students), and who have responsibilities to share effective instructional practices and/or to assess and improve the teaching effectiveness of other teachers in the school.”

The additional pay for serving as a master or lead teacher was, on average, larger than the pay-for-performance bonus. However, the pay-for-performance bonus was larger than all other opportunities for earning additional pay. In TIF districts that offered additional pay for serving as a master or lead teacher, the maximum amount offered was 33 percent more than the maximum amount offered for a pay-for-performance bonus. In TIF districts that offered additional pay for serving as a mentor teacher, the maximum amount offered was about 70 percent of the average maximum amount offered by all TIF districts for a pay-for-performance bonus. For assuming such responsibilities as serving as a department chair, participating in a schoolwide committee, or becoming a leadership team member, the maximum additional pay amount offered was about one-quarter of the maximum pay-for-performance bonus amount. We show the full set of results in Appendix D, Table D.3.

Districts could also choose to provide additional pay for working in hard-to-staff schools or subjects, but this was not a requirement of the grant. Twenty-four percent of TIF districts offered additional pay for working in a high-need subject area, and 17 percent offered additional pay for working in a hard-to-staff school. For districts offering these opportunities, the average maximum amount was \$3,455 for teaching in a high-need subject area and \$3,602 for teaching in a hard-to-staff school (Table III.4). Prior to TIF, fewer than 20 percent of TIF districts had given additional pay for teachers to teach in high-need subject areas (18 percent) or to work in hard-to-staff schools (13 percent) (Appendix D, Table D.2).

Requirement 4 -- Professional Development

The TIF notice required that applicants demonstrate a plan to provide high quality professional development that was directly linked to the measures of educator effectiveness. In other words, professional development would target individual teachers' and principals' needs, as identified in the evaluation process. The notice also required professional development to support educators' understanding and use of the measures of effectiveness. In this section, we describe TIF districts' reports about the focus of the professional development planned under TIF.

Most TIF districts planned to provide professional development to a majority of teachers. Seventy-one percent of TIF districts planned professional development that focused on providing teachers feedback based on their performance ratings (Table III.5). Eighty-seven percent of TIF districts planned to provide professional development to help educators understand the performance measures of the TIF program. Two-thirds of TIF districts had plans for professional development that focused on both of these topics (not shown).

Table III.5. Planned Focus of Professional Development (Percentages)

	All TIF Districts
Understanding performance measures of TIF program	87.0
Feedback based on TIF performance ratings	70.5
Number of Districts—Range^a	146–151

Source: District survey.

^aSample sizes are presented as a range based on the data available for each row in the table.

Implementing TIF Program Requirements

Taken together, the four required TIF program components constitute a comprehensive plan for a performance-based compensation system, and the grant required that all of the individual requirements be implemented together. In this section, we report on TIF grantees' success in implementing all components of the performance-based compensation system together.

Approximately seven percent of TIF grantees were unable to implement their proposed performance-based compensation system by the 2011–2012 school year and withdrew from the grant. As shown in Chapter II, 12 of the 183 TIF districts did not implement a TIF program in the 2011–2012 school year (see Table II.1). TIF grantees reported a variety of reasons for withdrawing from the grant, including (1) insufficient staff to execute the TIF program with fidelity, (2) inability to sustain support for the program among schools and staff, and (3) lack of union approval for either the proposed measure of educator effectiveness or bonuses based on the measures.²⁶

Fewer than half of districts reported implementing all four required program components for teachers. To examine whether TIF districts implemented all of the required program components, we first examined the percentage of districts that implemented all required elements for teachers.²⁷ Next, we measured the percentage of districts that reported all of the required elements except professional development. These components included: (1) using student achievement growth and at least two formal observations to measure educator effectiveness, (2) offering a pay-for-performance bonus, and (3) offering additional pay opportunities.²⁸ Forty-six percent of districts implemented all required components of TIF for teachers. When excluding the requirement for professional development, 68 percent of TIF districts implemented all of the other requirements for teachers, while 58 percent did so for principals.²⁹ However, half reported a TIF program that satisfied the TIF requirements—excluding professional development—for both teachers and principals (Table III.6).

Educator Involvement

Involving educators in planning and designing the TIF program may increase their support for and understanding of the program, and may promote greater satisfaction with the program. As discussed in Chapter I, this element was required prior to implementation to ensure strong implementation. In this section, we provide details about how districts involved educators in the TIF program design.

²⁶ All of the 12 districts that did not implement a TIF program in the 2011–2012 school year received a survey to complete for the study. Seven of them returned the survey, indicating that they withdrew from the TIF grant and providing brief explanations for their withdrawal.

²⁷ According to the original TIF notice, grantees could not use TIF program funds for incentive payments until they had implemented a performance-based compensation system that included all of the required components. Although most grantees used the 2010–2011 school year as a planning year, once grantees began implementation they were expected to implement all of the required components.

²⁸ Professional development for teachers and principals is also a requirement in the TIF notice. However, we excluded this requirement from some calculations because we did not have data for principals.

²⁹ The TIF notice required pay for additional opportunities for educators. Most grantees met this requirement by offering additional opportunities to teachers. Therefore, we did not include it as a requirement for principals.

Most TIF districts reported involving teachers or unions in designing their TIF programs, and used a combination of strategies to do so. We found that 95 percent of districts reported using this approach (Table III.7). Many districts used a combination of strategies, such as securing union approval and having teachers and principals vote directly on their district's proposed TIF program.

Table III.6. Implementation of TIF Program Requirements (Percentages)

	Teachers	Principals
TIF Requirements		
Requirement 1: Formal observations and student achievement growth ^a	79.3	68.0
Requirement 2: Pay-for-performance bonus	94.1	93.5
Requirement 3: Additional pay opportunities for teachers or principals	86.0 ^b	86.0 ^b
Requirement 4: Professional development	66.4	N/A ^c
Implemented all four requirements	46.0	N/A
Implemented all requirements except professional development	67.8	58.0
	Teachers and Principals	
Implemented all requirements except professional development	50.3	
Number of Districts—Range^d	137–152	150–153

^aTIF districts were required to use multiple formal observations for teachers and principals. For teachers, the survey asked whether districts used multiple formal observations. For principals, the survey asked only whether districts used formal observations, and did not ask about the number of observations.

^bThe 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.

^cWe do not have data on the percentage of districts that planned to provide professional development to principals.

^dSample sizes are presented as a range based on the data available for each row in the table.

Table III.7. Type of Educator Involvement in TIF Program Development (Percentages)

	All TIF Districts
Teacher and Union Involvement	
Any type of teacher or union involvement	94.7
Teachers' union voted on or approved TIF program	48.0
Teachers voted on or approved TIF program	62.4
Teacher groups served on formal design or planning committee	69.9
Principal Involvement	
Principals voted on or approved TIF program	57.0
Number of Districts—Range^a	148–153

Source: District survey.

^aSample sizes are presented as a range based on the data available for each row in the table.

Teachers may become involved in TIF program design if the teachers' union or association votes to approve the TIF program. Approval could occur through negotiations or collective bargaining, and is rarely achieved without also directly involving teachers in other ways. Among the 48 percent of TIF districts that had unions vote on or approve their program, nearly all of them (97 percent) also had teachers vote on or approve the program, or had a teacher representative on a formal design or planning committee.

Another approach to educator involvement was to have educators directly vote on or approve the TIF program (62 percent). States both with and without collective bargaining agreements used this method. However, almost all of the districts that held teacher votes but did not report union involvement were in right-to-work states (98 percent), where employees can decide whether to join or financially support a union. Finally, the Teacher Advancement Program (TAP), implemented by 37 percent of all TIF districts requires teacher votes before implementing the program, regardless of the districts' collective bargaining status.³⁰

TIF districts often combined approval from teachers or unions with involvement of teachers on a committee that designed or planned the TIF program. Seventy percent of TIF districts involved teachers on a design or planning committee. This involvement differs from teacher or union approval of the program, because the committee may give teachers an opportunity to provide feedback on the design of the TIF program. Among districts that obtained union or teacher approval for their TIF program, 70 percent also had teachers serve on a design or planning committee. Districts may have involved teachers on such committees to build teacher support for TIF before obtaining approval from a union or a broader group of teachers.

The involvement of educators may have led to different design decisions by TIF grantees. We examined whether TIF districts revised their programs after the grant award to obtain the support of educators. Most TIF districts (58 percent) reported revising their program from the original proposal in their grant application (Table III.8).³¹ Although 26 percent of TIF districts reported having made the changes to obtain the support of educators, the most frequently cited reason for revising TIF programs was to address budget limitations (31 percent). Seventeen percent of districts altered their programs after conducting an analysis of their educator performance measures to better predict future bonus payouts. Finally, 15 percent of districts revised their programs to better align with the districts' data systems, another core element.

Table III.8. Reasons Districts Reported for Revising Their Proposed TIF Program (Percentages)

	All TIF Districts
Percentage of Districts That Revised TIF Program After Grant Award	58.4
To address budget limitations	30.9
To obtain the support of educators	26.0
Based on results of analysis of educator performance measures	17.4
To better align with data management systems	14.9
Number of Districts—Range^a	146–149

Source: District survey.

^aSample sizes are presented as a range, based on the data available for each row in the table.

³⁰ To determine whether a TIF grantee district implemented the TAP, we reviewed TIF profiles from the Center for Education Compensation Reform website.

³¹ Although we did not ask about the extent of revisions to their TIF programs, grantees were required to implement a TIF program that was consistent with their original grant application.

IV. IMPLEMENTATION OF TIF IN THE EVALUATION DISTRICTS

In this chapter, we describe the implementation of TIF by the subset of grantees awarded a grant to participate in a random assignment evaluation of the pay-for-performance component of TIF. The first section is based on all 12 evaluation districts; the next two sections exclude the 2 districts that were not prepared for random assignment until the end of the 2011–2012 school year.

We focus on whether evaluation districts implemented the four required program components described in Chapter I. These program components—particularly, measures of educator effectiveness and the planned pay-for-performance bonuses—could influence educator behaviors and, ultimately, student achievement. We also provide information about how evaluation districts implemented these requirements, using details we obtained from evaluation districts through phone interviews and technical assistance documents. In the second part of this chapter, we examine teachers’ and principals’ understanding of the TIF program in their districts. Educators can be motivated to change their teaching practices or change schools in response to bonus offers only if they understand how they are being evaluated and how much they may potentially earn. It is important, therefore, to understand not only how districts structured their evaluation systems and bonuses but also the degree to which educators understood them. We used surveys of principals and teachers, administered only in evaluation districts, to measure their understanding of the TIF program.

Key Findings on TIF Implementation in Evaluation Districts

- **About three-quarters of the evaluation districts implemented all of the required components of TIF for teachers.**
- **Three-quarters of evaluation districts implemented the required components of TIF related to measures of educator effectiveness and pay-for-performance for principals.** One-quarter of the evaluation districts did not conduct observations of principals with trained observers.
- **Consistent with the TIF grant goals, evaluation grantees expected pay-for-performance bonuses to be somewhat substantial and differentiated. However, the districts expected most educators would receive a bonus, suggesting that the award criteria were not consistent with TIF guidance for challenging pay-for-performance bonuses.**
- **Evaluation districts offered separate bonuses for different performance measures, and offered teachers larger bonuses for performance based on student achievement growth than for classroom observations.**
- **In evaluation districts, educators’ reported awareness of performance measures often differed from districts’ reports; principals’ reports were more consistent with districts’ reports.**
- **Teachers and principals in treatment schools reported lower rates of eligibility for pay-for-performance bonuses and lower expected pay-for-performance bonuses than districts reported.**

Design of TIF Programs in Evaluation Districts

In this section, we examine the design of TIF programs in evaluation districts, focusing 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. Understanding the design and implementation of the TIF grant in evaluation districts is important for informing its impact on educators' attitudes and behavior. Future reports will study its effect on student achievement and educator mobility.

We also compare required program components in evaluation and non-evaluation districts to provide additional context for the study's findings. This approach can inform whether the performance-based compensation systems implemented by TIF evaluation districts are similar to those implemented by all TIF districts. The TIF programs designed by evaluation and non-evaluation districts may have differed for two reasons. First, the TIF notice gave evaluation grantees additional guidance on designing their pay-for-performance bonuses, specifying that bonuses should be substantial, challenging, and differentiated. This guidance, including specific examples, may have led to differences in the pay-for-performance bonuses designed by evaluation and non-evaluation districts. Second, differences in the characteristics of evaluation and non-evaluation districts, and, therefore, their local context, may have influenced decisions about the TIF program design. As noted in Chapter II, evaluation districts were often larger, more urban, and located in states with collective bargaining.

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 rewarding teachers and principals with performance-based bonuses.

All evaluation districts reported using student achievement growth to evaluate teachers and principals (Table IV.1). Although the grant notice required them to use achievement growth to evaluate teachers and principals, districts had flexibility in choosing which growth measure they used. All of the evaluation districts used achievement growth for the entire school to evaluate teachers, and two-thirds of evaluation districts also used achievement growth of students in teachers' own classrooms. Forty-six percent of evaluation districts used achievement growth based on groups of teachers. For principals, all evaluation districts used achievement growth for the entire school, and two-thirds used growth measures based on groups of teachers.

Evaluation and non-evaluation districts differed in their use of achievement growth to evaluate teachers in TIF schools. Despite grant requirements, 18 percent of non-evaluation districts did not report using student achievement growth to evaluate teachers, and 11 percent of non-evaluation districts did not report using it to evaluate principals (Table IV.1). Evaluation districts were more likely to report using achievement growth for the entire school (100 percent of evaluation districts, versus 74 percent for teachers and 89 percent for principals in non-evaluation districts). Differences in the percentage of evaluation and non-evaluation districts assessing teachers and principals on student achievement levels (that is, achievement at one point in time, such as the percentage of students scoring "proficient" on a standardized test) were not statistically significant.

Table IV.1. Percentage of Districts Using Student Achievement and Observation Measures for Teachers and Principals, by Evaluation Participation Status

Teacher Performance Measure	Teachers		Principals	
	Evaluation Districts	Non-Evaluation Districts	Evaluation Districts	Non-Evaluation Districts
Student Achievement Measures				
Any achievement measure	100.0*	81.9	100.0	90.8
Achievement growth	100.0*	81.9	100.0*	89.4
By schools	100.0*	73.9	100.0*	89.4
By student subgroups ^a	45.5	48.6	66.7*	51.1
By teachers' classrooms	66.7	69.6	-	-
Achievement level	41.7	45.7	66.7	65.2
Observation Measures				
Classroom observations	100.0	96.4	-	-
At least two classroom observations	100.0*	94.8	-	-
Principal observations	-	-	75.0	74.5
Number of Districts—Range^b	11–12	135–138	12	139–141

Source: District survey.

^aStudent subgroups can be defined by grade, teams, subject areas, and demographic characteristics.

^bSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

All evaluation districts reported using formal observations to evaluate teachers, and 9 of the 12 districts reported using observations for principals (Table IV.1). All of the evaluation districts reported using at least two formal classroom observations to evaluate teachers in TIF schools, as required by the grant notice. Evaluation districts planned to conduct three observations per teacher, with the average observation lasting about one class period, or 40 minutes (Appendix D, Table D.4). Most evaluation districts used principals as well as other staff, such as teacher leaders, peer observers, or content specialists, to conduct observations of teachers.

Evaluation and non-evaluation districts were similar in their use of formal observations to evaluate teachers or principals. Although evaluation districts were statistically more likely to conduct at least two classroom observations for each teacher, the difference was relatively small (100 percent of evaluation districts versus 95 percent of non-evaluation districts). We found no statistically significant differences between evaluation and non-evaluation districts in the number of observations per school year, the length of observations, or the types of staff conducting observations (Appendix D, Table D.4). Despite a grant requirement, only three-quarters of both evaluation and non-evaluation districts reported using principal observations conducted by trained observers (Table IV.1).

Requirement 2 -- Pay-for-Performance Bonuses

In this section, we describe the pay-for-performance bonuses offered by evaluation districts. This component is the only one that was designed to differ across treatment and control schools. In addition to reporting the percentage of evaluation districts that met the requirement to offer pay-for-performance bonuses, we also examine the expected size of pay-for-performance bonuses and how districts expected them to be distributed across educators. These aspects of the pay-for-performance bonuses can determine their influence on teacher behavior, and, ultimately, student outcomes. More

specifically, we focus on whether evaluation districts expected to distribute pay-for-performance bonuses that aligned with TIF guidance that bonuses be *substantial, challenging, and differentiated*. Any differences in the design of the pay-for-performance bonus in evaluation and non-evaluation districts may be because only the evaluation grantees received additional guidance on the meaning of those terms.

Throughout this section, we report on the maximum and minimum bonuses districts expected to offer teachers and principals. We also note the average bonus that districts expected to pay to educators. The amount is “expected,” because districts estimated the percentage of educators who would earn different bonus amounts (for example, the percentage of educators expected to earn a bonus of \$1,000 to \$1,999). Although all districts had to estimate how teacher performance would be distributed, the availability of information to inform these estimates varied. For example, districts using existing performance measures for their TIF program would have more information on the distribution of teacher performance than districts that were using new measures and may have had to guess what percentage of educators would earn each bonus amount. As noted in Chapter III, many districts were unable to respond to this question, resulting in lower sample sizes.

All evaluation districts reported offering pay-for-performance bonuses. This finding differed in the non-evaluation districts, of which 94 percent reported offering pay-for-performance bonuses for teachers and 93 percent reported offering them for principals.³²

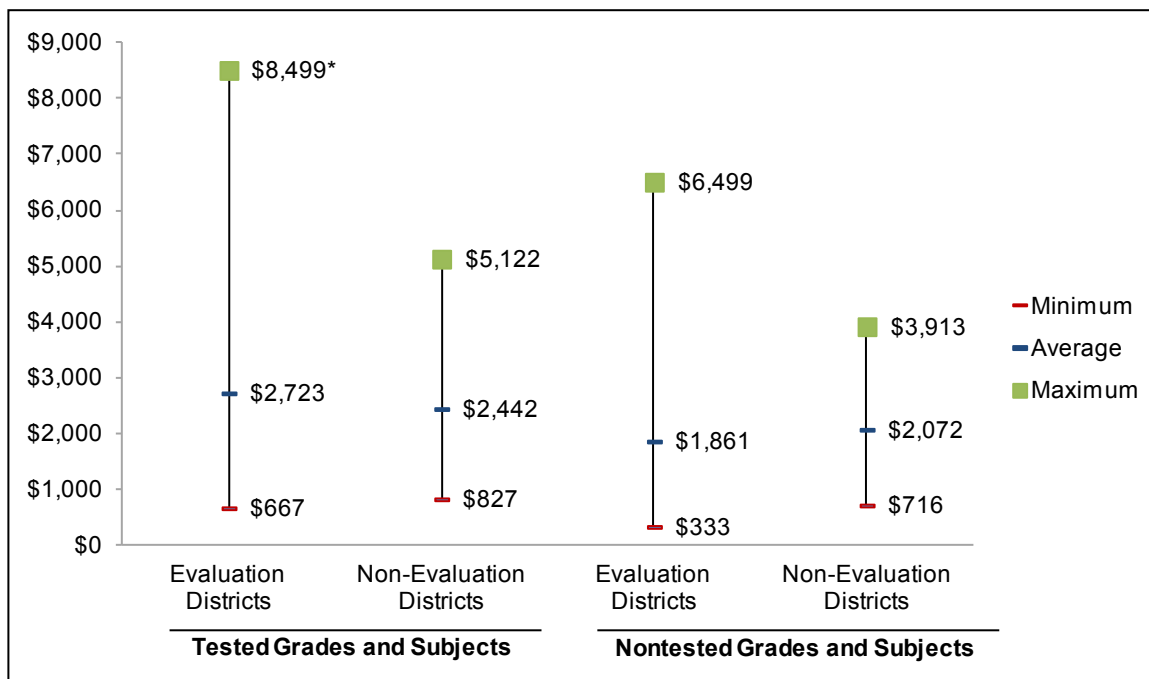
Evaluation districts expected pay-for-performance bonuses to be somewhat substantial and differentiated. However, the districts expected most educators would receive a bonus, suggesting that the award criteria were not consistent with TIF guidance for challenging pay-for-performance bonuses. We compared evaluation and non-evaluation districts on whether their pay-for-performance bonuses met the criteria suggested in the grant notice: that the bonuses be (1) substantial, (2) challenging, and (3) differentiated.

Evaluation districts expected average pay-for-performance bonuses that were very close to the 5 percent of average salary provided as an example in the TIF notice. In Figure IV.1, we show that evaluation districts offered a maximum pay-for-performance bonus of \$8,499 for teachers in tested grades and subjects, and an expected average pay-for-performance bonus of \$2,723. The average bonuses expected by evaluation districts were 4.8 percent of average salary.³³ Evaluation districts offered smaller pay-for-performance bonuses for teachers in nontested grades and subjects: a maximum pay-for-performance bonus of \$6,499 and an average bonus of \$1,861 (or 3.3 percent of average salary). The \$9,571 average expected pay-for-performance bonus for principals represented 4.0 percent of the average U.S. principal salary. Although evaluation districts offered a larger maximum pay-for-performance bonus than non-evaluation districts, the difference between evaluation and non-evaluation districts in the expected average pay-for-performance bonus was not statistically significant. In addition, there were no statistically significant differences across evaluation and non-evaluation districts in the maximum, minimum, or average pay-for-performance bonuses offered for teachers in nontested grades and subjects or for principals (Figure IV.2).

³² Differences between evaluation and non-evaluation districts in the percentage offering pay-for-performance bonuses for teachers and principals were significant at the 0.05 level.

³³ These percentages are based on the average U.S. teacher’s salary, as reported in the U.S. Department of Education’s School and Staffing Survey (http://nces.ed.gov/programs/digest/d12/tables/dt12_083.asp).

Figure IV.1. Expected Pay-for-Performance Bonuses for Teachers in Evaluation and Non-Evaluation Districts, Averages Across Districts



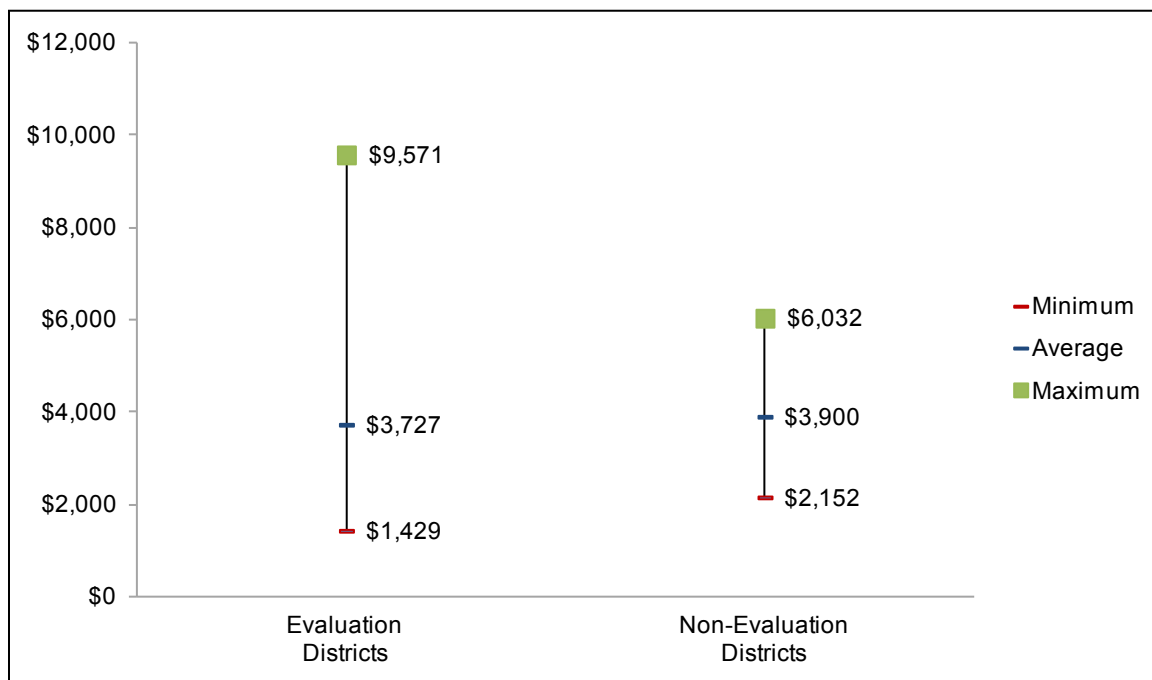
Source: District survey.

Note: Figures are based on survey question about the expected distribution of TIF-funded pay-for-performance bonuses, given 10 categories of bonus amounts that range from \$0 to \$15,000 or more (for example, the percentage of teachers expected to earn a bonus between \$1,000 and \$1,999). Six of the 12 evaluation districts and 81 of 141 the non-evaluation districts responded to this question. For each district, the maximum bonus was calculated as the upper range of the largest category with teachers expected to receive a bonus in that range; the minimum bonus was based on the lower range of the lowest category with teachers expected to receive a bonus in that range; and the average bonus was calculated as the average of the midpoint dollar amount of each category, weighted by the percentage of teachers expected to receive a bonus in that range.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

Evaluation districts expected that most teachers and principals would earn a pay-for-performance bonus, suggesting it was not aligned with the TIF guidance for challenging pay-for-performance bonuses. We examined the percentage of teachers not expected to earn a bonus to assess whether evaluation districts thought obtaining pay-for-performance bonuses would be challenging. In Figures IV.3 and IV.4, we show the percentage of educators expected to earn a pay-for-performance bonus at various amounts. Evaluation districts expected that 76 percent of teachers and principals would earn pay-for-performance bonuses. The TIF evaluation competition notice suggested that pay-for-performance bonuses should be given to educators who perform “significantly better than the current average performance” among schools in a district. Non-evaluation districts expected a significantly higher percentage of teachers and principals to earn a pay-for-performance bonus: 94 percent of teachers and 97 percent of principals.

Figure IV.2. Expected Pay-for-Performance Bonuses for Principals in Evaluation and Non-Evaluation Districts, Averages Across Districts



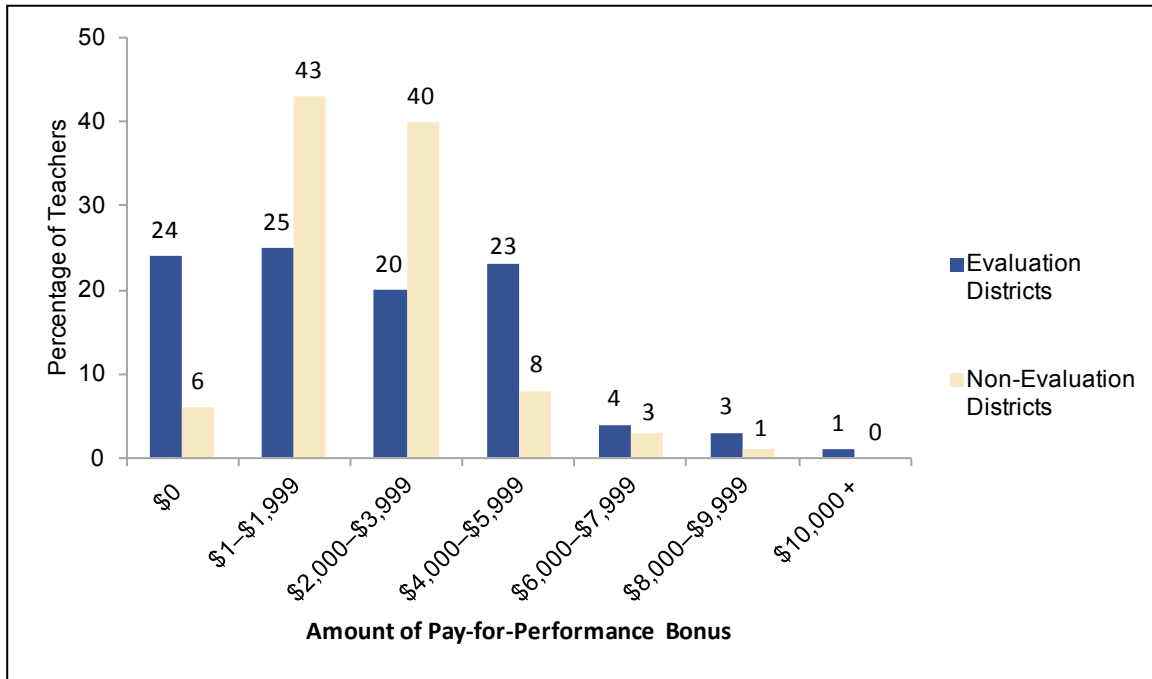
Source: District survey.

Note: Figures are based on survey question about the expected distribution of TIF pay-for-performance bonuses, given 10 categories of bonus amounts that range from \$0 to \$15,000 or more (for example, the percentage of principals expected to earn a bonus between \$1,000 and \$1,999). Seven of the 12 evaluation districts and 92 of the 141 non-evaluation districts responded to this question. For each district, the maximum bonus was calculated as the upper range of the largest category with principals expected to receive a bonus in that range; the minimum bonus was based on the lower range of the lowest category with principals expected to receive a bonus in that range; and the average bonus was calculated as the average of the midpoint dollar amount of each category, weighted by the percentage of principals expected to receive a bonus in that range.

Differences in the minimum, maximum, and average pay-for-performance bonus between evaluation and non-evaluation districts were not statistically significant at the 0.05 level, two-tailed test.

Evaluation districts expected pay-for-performance bonuses for teachers that aligned with the TIF guidance for differentiated bonuses. Evaluation districts expected to pay teachers a maximum pay-for-performance bonus that was three times the expected average bonus (maximum bonus of \$8,499 is 3.1 times larger than the average bonus of \$2,723). Evaluation districts expected maximum pay-for-performance bonuses for principals that were 2.6 times the average bonus, a bit lower than the TIF notice suggested. In Figures IV.3 and IV.4, we show the distribution of pay-for-performance bonuses expected by districts for teachers and principals, respectively. Evaluation districts expected 45 percent of teachers and 40 percent of principals to earn a pay-for-performance bonus of \$1 to \$3,999. Evaluation districts expected more differentiated pay-for-performance bonuses for teachers and principals than non-evaluation districts. Non-evaluation districts expected to pay teachers a maximum bonus that was double the average bonus (maximum bonus of \$5,122 is 2.1 times larger than the average bonus of \$2,442), and principals a maximum pay-for-performance bonus that was 1.5 times the average. In addition, non-evaluation districts expected 83 percent of teachers and 61 percent of principals to earn pay-for-performance bonuses of \$1 to \$3,999 (Figures IV.3 and IV.4).

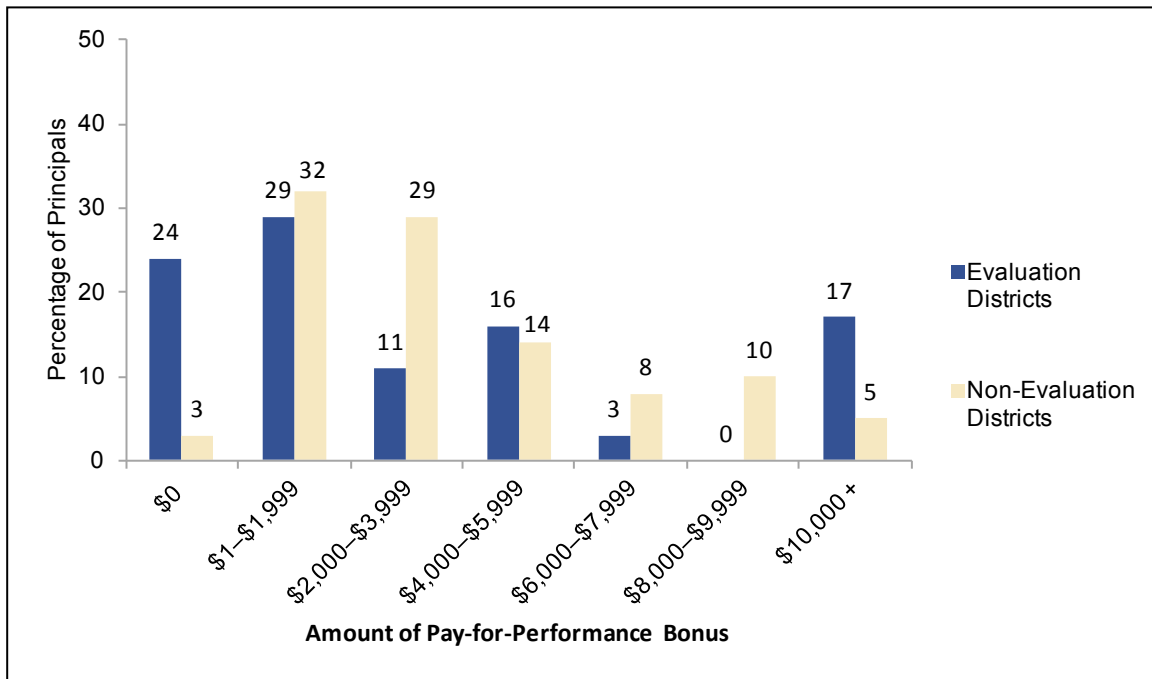
Figure IV.3. Distribution of Expected Pay-for-Performance Bonuses for Teachers in Tested Grades and Subjects, by Evaluation Participation Status



Source: District survey.

Note: Six evaluation and 81 non-evaluation districts responded to this survey question.

Figure IV.4. Distribution of Expected Pay-for-Performance Bonuses for Principals, by Evaluation Participation Status



Source: District survey.

Note: Seven evaluation and 92 non-evaluation districts responded to this survey question.

Requirement 3 -- Additional Pay Opportunities

The TIF grant required that districts provide additional pay for effective teachers to take on additional roles and responsibilities. Examples from the TIF notice include serving as a master or mentor teacher to help other teachers improve, mentoring novice teachers, tutoring students, and developing learning communities. We examined the percentage of evaluation districts that provided these additional pay opportunities, the types of roles and responsibilities offered, and the size of the pay opportunities.

All evaluation districts offered additional pay opportunities for teachers (Table IV.2).

The additional pay opportunities most commonly offered by evaluation districts were for mentor teachers (91 percent of evaluation districts) and for master or lead teachers (73 percent of evaluation districts). A higher percentage of evaluation districts than non-evaluation districts offered additional pay for teachers to assume additional responsibilities (100 percent of evaluation districts versus 86 percent of non-evaluation districts).

We compared the amount of money teachers could earn for these additional pay opportunities with the maximum amount they could earn for pay-for-performance bonuses. This comparison is particularly relevant for future reports on the eventual impact of the pay-for-performance component of the performance-based compensation system, because teachers in treatment and control schools were eligible for these additional pay opportunities. Because districts may have selected teachers for these roles and responsibilities based in part on their effectiveness, an additional pay opportunity equal to a pay-for-performance bonus may create similar incentives for teachers in treatment and control schools, thereby diminishing the effect of offering pay-for-performance. However, additional pay opportunities may 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.

For evaluation districts, the average maximum additional pay of \$3,460 for serving as a mentor teacher (among evaluation districts offering this type of pay) represented 41 percent of the average maximum pay-for-performance bonus of \$8,499 (among all evaluation districts). The maximum pay of \$5,104 for serving as a master or lead teacher in evaluation districts offering this type of pay represented 60 percent of the \$8,499 average maximum pay-for-performance bonus for all evaluation districts (Appendix D, Table D.6).

Requirement 4 -- Professional Development

The TIF grant required that districts provide professional development linked to the educator evaluation measures. This support included professional development to help educators understand the measures being used to evaluate their performance as well as feedback based on evaluation ratings to help educators improve their instructional practices. We asked evaluation and non-evaluation districts whether they had planned professional development for teachers in TIF schools that focused on these two topics. A similar percentage of evaluation and non-evaluation districts (91.7 and 86.6 percent, respectively) planned to provide professional development focused on each of these topics; there were no statistically significant differences between the two types of districts (Table IV.3). Additionally, two-thirds of evaluation and non-evaluation districts planned professional development on both topics through the TIF program (not shown).

Table IV.2. Additional Pay Opportunities for Teachers, Comparison of TIF Evaluation and Non-Evaluation Districts

	Evaluation Districts		Non-Evaluation Districts	
	Percentage That Offered Additional Pay	Maximum Pay in Districts Offering Additional Pay	Percentage That Offered Additional Pay	Maximum Pay in Districts Offering Additional Pay
Teachers could receive additional pay for taking on added roles or responsibilities	100.0*	n.a.	85.5	n.a.
Roles and Responsibilities				
Mentor teacher	90.9*	\$3,460	64.2	\$3,770
Master or lead teacher	72.7	\$5,104	53.7	\$7,400
Support school-, grade-, or subject-level decisions ^a	50.0	\$2,542	39.0	\$1,495
Additional Factors				
Teaching in hard-to-staff school or in high-need subject areas	33.3	\$4,725	30.3	\$3,518
Attending professional development activities or enrolling in graduate-level courses	33.3	\$633	27.3	\$796
Number of Districts—Range^b	11–12	3–10	132–141	28–78

Source: District survey.

Note: Table reports on activities funded by TIF.

^aIncludes being a department chair or a lead curriculum specialist, or serving on a schoolwide committee or leadership team.

^bSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

n.a. = not applicable.

Table IV.3. Percentage of Districts Reporting Professional Development Activities for Teachers Planned Under TIF, by Evaluation Participation Status

	Evaluation Districts	Non-Evaluation Districts
Focus of Professional Development		
Understanding performance measures used for TIF	91.7	86.6
Feedback to teachers based on TIF performance ratings	66.7	70.9
Number of Districts—Range^a	11–12	134–140

Source: District survey.

Note: Differences between evaluation and non-evaluation districts are not statistically significant at the 0.05 level.

^aSample sizes are presented as a range based on the data available for each row in the table.

Implementing Multiple TIF Program Requirements

In addition to examining the percentage of evaluation districts implementing each TIF requirement separately, we also measured the percentage of districts that implemented all of the requirements together. First, we examined whether TIF evaluation districts implemented all four of the required components of TIF for teachers. Next, we examined whether evaluation districts implemented all required components except for professional development for teachers and principals. We excluded the requirement for professional development because we did not have data for principals. Because we collected information from districts on planned professional development in the middle of the first year of implementation, many districts may not have yet had the opportunity to inform educators about, for example, their performance in terms of student achievement growth. For principals, we considered whether districts measured principal effectiveness based on student achievement growth and observations, and whether they offered a pay-for-performance bonus.

About three-quarters of the evaluation districts implemented all four of the required components of TIF for teachers (Table IV.4). Excluding professional development, all evaluation districts implemented the required components of TIF for teachers, while 73 percent of evaluation districts did so for principals. For principals, all of the evaluation districts used student achievement growth to measure performance and offered pay-for-performance bonuses, but only 75 percent conducted observations of principals with trained observers (Table IV.1). Evaluation districts were more likely than non-evaluation districts to implement all of the required components of TIF for teachers (73 versus 44 percent), and all required component except professional development for teachers (100 versus 65 percent) and principals (73 versus 57 percent).

Educator Involvement

In addition to the TIF program requirements described above, TIF districts also had to implement five core elements to support their performance-based compensation systems. In this section, we describe how districts implemented one of those core elements: involving educators in the design of TIF programs. The involvement of educators may influence the extent to which educators support and understand the TIF programs.

All evaluation districts involved teachers or teachers' unions in the development of their TIF programs (Table IV.5). Two-thirds of evaluation districts had teachers participate on a design or planning committee, and 55 percent of the districts reported holding a vote or obtaining approval from teachers or teachers' unions. Fewer evaluation districts (45 percent) had principals vote on or approve the TIF program. Evaluation districts were statistically more likely than non-evaluation districts to involve teachers in the development of their TIF programs, but the difference was small (100 percent of evaluation districts versus 94 percent of non-evaluation districts). Evaluation and non-evaluation districts did not differ significantly in how they involved teachers and unions. Furthermore, there were no statistically significant differences in the percentages of evaluation and non-evaluation districts that involved principals in the development of their TIF program.

Table IV.4. Percentage of Districts Reporting Implementation of TIF Program Requirements, by Evaluation Participation Status

	Teachers		Principals	
	Evaluation Districts	Non-Evaluation Districts	Evaluation Districts	Non-Evaluation Districts
TIF Requirements				
Requirement 1: At least two formal observations and student achievement growth ^a	100.0*	77.4	75.0	67.4
Requirement 2: Pay-for-performance bonus	100.0*	93.6	100.0*	92.9
Requirement 3: Additional pay opportunities for teachers or principals ^b	100.0*	84.9	100.0*	84.9
Requirement 4: Professional development	66.7	64.4	N/A ^c	N/A ^c
Implemented all four requirements	72.7*	43.7	N/A	N/A
Implemented all requirements except professional development	100.0*	65.2	72.7	56.8
Teachers and Principals				
	Evaluation Districts		Non-Evaluation Districts	
Implemented all requirements except professional development	72.7		48.5	
Number of Districts—Range^d	11-12	126-140	11-12	139-141

^aTIF districts were required to use multiple formal observations for teachers and principals. For teachers, we have information on whether districts used multiple formal observations. For principals, we have information on whether districts used formal observations (without considering the number of observations).

^bThe TIF grant notice required that districts provide additional pay opportunities for educators, so these percentages are based on the percentage of TIF districts offering these pay opportunities to teachers or principals.

^cWe do not have data on the percentage of districts that planned to provide professional development for principals.

^dSample sizes are presented as a range based on the data available for each row in the table.

*Differences between evaluation and non-evaluation districts are statistically significant at the 0.05 level.

Table IV.5. Educator Involvement in TIF Program Development (percentages), by Evaluation Participation Status

	Evaluation Districts	Non-Evaluation Districts
Teacher and Union Involvement		
Any type of teacher or union involvement	100.0*	94.3
Teachers' union voted on or approved TIF program	54.5	47.4
Teachers voted on or approved TIF program	54.5	63.0
Teacher groups served on formal design or planning committee	66.7	70.2
Principal Involvement		
Principals voted on or approved TIF program	45.5	58.0
Number of Districts—Range^a	11-12	137-141

Source: District survey.

Note: Sample size may vary for individual items due to item nonresponse. The table shows the minimum and maximum sample sizes.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

The involvement of teachers may have influenced how TIF districts designed their programs. We examined whether districts revised their programs to obtain the support of educators. We focused on revisions that TIF districts made to their programs after the grant award—during the planning year, for districts that had not yet met the five core elements for supporting implementation. Eighty-two percent of evaluation districts revised their programs after the grant award (Table IV.6), including evaluation districts that made revisions to obtain educator support (55 percent) and to address budget limitations (55 percent). A smaller percentage of evaluation districts made revisions to accommodate data management systems (18 percent) or in response to an analysis of their proposed evaluation measures (9 percent). Fewer non-evaluation districts (57 percent) revised their TIF programs after the grant award compared with evaluation districts. In addition, evaluation districts were more likely than non-evaluation districts to report making revisions to obtain the support of educators (24 percent of non-evaluation districts). A similar percentage of evaluation and non-evaluation districts cited the other reasons for revising their programs.

Table IV.6. Reasons Districts Reported for Revising Their Proposed TIF Programs, by Evaluation Participation Status (percentages)

	Evaluation Districts	Non-Evaluation Districts
Revised TIF program after grant award	81.8*	56.5
To conform to/address budget limitations	54.5	29.0
To obtain the support of educators	54.5*	23.7
Based on results of an analysis of proposed bonus system	9.1	18.1
To accommodate data management systems	18.2	14.6
Number of Districts—Range^a	11	135–138

Source: District survey.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

Implementation of Measures of Educator Effectiveness and Pay-for-Performance Bonuses in Evaluation Districts

Although all TIF districts had to meet the same program requirements described above, they had some discretion in how they met these requirements. We use additional information on evaluation grantees from telephone interviews and technical assistance documents to provide insight into how districts developed measures of educator effectiveness based on student achievement growth and multiple classroom observations (characterized as performance measures below). We also look at how they linked these measures to pay-for-performance bonuses. We first describe implementation of the educator performance measures required by TIF, and then examine how evaluation districts used these measures to determine pay-for-performance bonuses. This section describes implementation findings for the 10 evaluation districts that were included in the random assignment study design for the 2011–2012 school year.

Implementation of Teacher and Principal Performance Measures in Evaluation Districts

The performance measures used to evaluate teachers and principals, and ultimately determine pay-for-performance bonuses, are a critical component of TIF programs. Given this importance, we examined how evaluation districts implemented performance measures for teachers and principals so we could provide additional context and insight into the potential challenges that TIF districts experienced.

All evaluation districts used a vendor in some capacity to develop student achievement growth measures. An important decision when using achievement growth to evaluate teachers and principals is whether to use an existing measure or develop a new one. Eight of the 10 evaluation districts used for their performance-based compensation system achievement growth measures obtained from the state or directly from a vendor (for example, Education Value-Added Assessment System or the Colorado Growth Model). The remaining two districts developed their own student achievement growth measures with assistance from a vendor. In each of these districts, a team of teachers and district staff had responsibility for making key decisions about the measure's design.

Most evaluation districts used new observation rubrics. TIF required the use of an objective, evidenced-based rubric when conducting observations of teachers and principals. Seven of the 10 evaluation districts used a classroom observation rubric that had not been used in the district prior to TIF. Among these seven districts, three implemented the Teacher Advancement Program (TAP) and used its observation rubric; the other four used or adapted an existing standards-based rubric (for example, the Danielson Framework). The remaining three districts relied on observation rubrics for TIF that were already in use: one district used a state-developed tool, and two districts used a rubric based on the Danielson framework. Among the seven evaluation districts that reported using a principal observation rubric, five used a principal observation tool new to the district. Four of these five districts developed or identified a new observation rubric and added it to an existing principal evaluation system or combined it with other principal measures, such as the Vanderbilt Assessment for Leadership in Education. The fifth district used a new tool provided by its state. The remaining two districts used an existing principal evaluation system developed by their states.

A majority of evaluation districts assessed the reliability of classroom observers. The TIF notice required that districts ensure a high degree of agreement among observers in their ratings of teachers and principals based on the observation rubrics. Given that neither the TIF notice nor subsequent guidance from ED defined how grantees should achieve this agreement, grantees used different approaches. For example, some grantees compared observer ratings with ratings from an observer whose assessments were considered the gold standard, and others compared observers with one another. Although grantees may have conducted additional checks of reliability after the observations began, we focused on how grantees checked reliability before observations began. Six of the 10 evaluation districts formally assessed the reliability of staff observing teachers by requiring that observers pass a certification test, comparing observers' ratings with those of an expert observer, or measuring the level of agreement among all observers. The remaining four districts did not formally assess the reliability of staff observing teachers.³⁴

Use of Performance Measures for Pay-for-Performance Bonuses in Evaluation Districts

Although TIF required that districts use achievement growth and multiple classroom observations to evaluate teachers and principals, districts decided how to use these measures for their pay-for-performance bonuses. We examined whether evaluation districts offered a single pay-for-performance bonus based on multiple performance measures or offered a separate bonus for each performance measure. We also focused on the relative weight given to measures based on student achievement growth, achievement level, and classroom observations. In examining student achievement growth, we identified the weight given to achievement growth measures based on

³⁴ We focus on whether grantees measured the reliability of classroom observers but not of principal observers, because we did not have sufficient information from districts about their efforts to assess the reliability of principal observers.

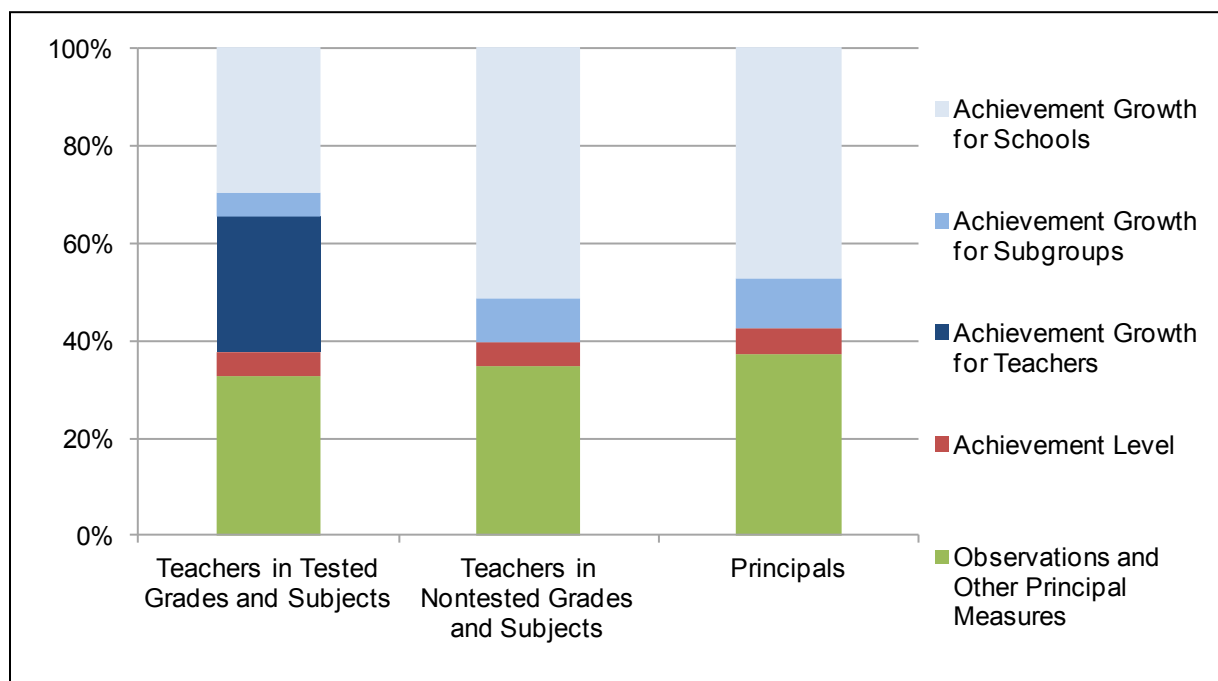
individual teachers' classrooms, groups of teachers, and entire schools. Differences among districts in the use of performance measures may reflect contextual factors, such as the availability of data to measure achievement for individual teachers, the preferences of teachers and principals involved in the design of the TIF program, or a district's past experience with pay-for-performance bonuses. The use of performance measures for pay-for-performance bonuses in evaluation districts is important context for the study's impact findings.

Most evaluation districts awarded a separate pay-for-performance bonus for each evaluation measure. Eight of the 10 evaluation districts offered a separate bonus for each type of achievement growth measure (for example, one bonus for student achievement growth for the entire school and one for student achievement growth in individual teachers' classrooms) and a separate bonus based on observations of teachers in their classrooms. The two remaining districts used a classroom observation measure to determine whether teachers were eligible for a bonus that was based on achievement growth. For example, all teachers earning a satisfactory rating on the classroom observation measure could earn a pay-for-performance bonus based on student achievement growth.

Evaluation districts offered larger bonuses for student achievement growth than for classroom observations. The TIF notice required that districts use achievement growth as a significant factor in evaluating teachers and principals. We examined how evaluation districts prioritized the various measures for their TIF program by comparing the relative size of bonuses for each type of measure. In Figure IV.5, we show that, on average, bonuses based on achievement growth made up more than half of the expected total bonus for teachers and principals. For teachers in tested grades and subjects, achievement growth accounted for 62 percent of the bonus, classroom observations were 33 percent, and measures based on student achievement level made up 5 percent. The relative sizes of bonuses for each measure were similar for teachers of nontested grades (61 percent for achievement growth, 35 percent for observations, and 5 percent for student achievement level), and for principals (55 percent, 39 percent, and 5 percent, respectively).

For teachers in grades and subjects that were not tested annually, all of the districts that used growth for individual teachers in tested grades and subjects assigned a greater weight to student achievement growth for the entire school, to evaluate these teachers as well as principals. School achievement growth accounted for 30 percent of the bonuses for teachers in tested grades and subjects, compared to 52 percent for nontested grades and subjects, and 47 percent for principals.

Student achievement growth in individual teachers' classrooms accounted for varying amounts of the performance bonus. A pay-for-performance bonus may incentivize certain behaviors depending on whether performance is measured for individual teachers, for a subgroup of teachers, or for an entire school. For example, a school-based measure might encourage teachers within a school to collaborate or share resources, or it might discourage this collaboration among teachers who feel they cannot influence the performance of other teachers in the school. A growth measure for individual teachers may prompt teachers to focus exclusively on performance in their own classrooms and could also discourage collaboration.

Figure IV.5. Relative Weight of Each Type of Performance Measure Used for Pay-for-Performance Bonuses in TIF Evaluation Districts

Source: Technical assistance documents.

Notes: Ten evaluation districts. Because some evaluation districts combined a principal observation measure with other measures, such as surveys of teachers and parents, we combine these measures into one category for principals.

Given the potential importance of the type of achievement growth measure used for bonuses, we examined the relative size of reported bonuses for each type of measure for teachers in tested grades and subjects (Figure IV.6). We grouped evaluation districts into three categories according to the relative weight of the bonus that is based on achievement growth for individual teachers:³⁵

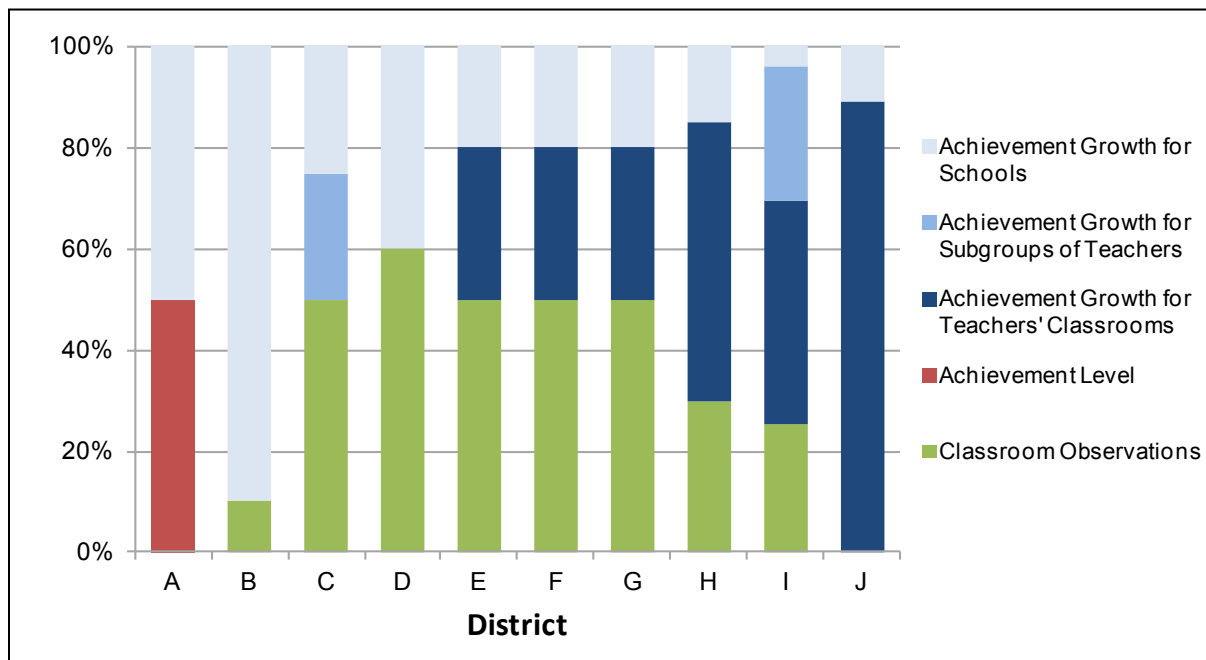
1. **Districts that do not use student achievement growth of individual teachers' classrooms to determine pay-for-performance bonuses.** Four evaluation districts did not evaluate teachers based on student achievement growth of individual teachers' classrooms (Figure IV.6, Districts A through D). Instead, these districts assigned greater weight to achievement growth for the entire school and for subgroups of teachers. District A combined school achievement growth with schoolwide measures based on student achievement levels. The relative size of bonuses based on classroom observations varied across these districts, from zero percent (a district that used classroom observations as an eligibility criterion to earn a bonus) to 60 percent.
2. **Districts that use student achievement growth in individual teachers' classrooms to determine no more than one-third of pay-for-performance bonus.** The three districts implementing TAP used student achievement growth in individual teachers' classrooms for 30 percent of the total bonus (Figure IV.6, Districts E through G). TAP

³⁵ We do not have a similar grouping for non-evaluation districts. The data used to develop these groupings came from a combination of the survey administered to all TIF districts administrators, responses to phone interviews with TIF administrators, and technical assistance documents. Information from the interviews and technical assistance documents were only collected in the evaluation districts.

specifies weights for each type of measure that districts can use or adapt. All three evaluation districts used the suggested weights provided by TAP: 20 percent of the bonus for student achievement growth for the entire school and 50 percent for classroom observations.

3. **Districts that use student achievement growth in individual teachers’ classrooms to determine more than one-third of pay-for-performance bonus.** In the remaining three districts, bonuses based on student achievement growth in individual teachers’ classrooms accounted for more than one-third of the pay-for-performance bonus (Figure IV.6, Districts H through J). Achievement growth for individual teachers accounted for 89 percent of the bonus in District J, which used classroom observations as an eligibility criterion. Measures based on classroom observations accounted for 30 percent or less of the pay-for-performance bonus in the other two districts.

Figure IV.6. Relative Weight of Each Type of Measure Used for Performance Bonuses for Teachers in Tested Grades and Subjects in Evaluation Districts



Source: Technical assistance documents.

Note: Ten evaluation districts.

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 will ultimately influence the educators’ behaviors. Moreover, educators’ reports about program features can identify ways in which their understanding of the TIF program deviates from what grantees intended or what district officials reported, highlighting possible challenges in the implementation process.

Treatment and control schools were expected to implement the same components of the district’s performance-based compensation system except for one component—pay-for-performance. The teacher and principal surveys included questions about eligibility for and potential magnitude of pay-for-performance bonuses. We compared responses from educators in treatment

and control schools to measure how consistent educators' beliefs were with the payout available at their school. For all other required components of TIF, we focused less on treatment-control differences and examined instead the degree to which teacher and principal reports were consistent with each other and with the description offered by district officials.³⁶ These analyses reveal how knowledgeable educators were about key program features. They can also identify possible discrepancies between intended and actual implementation of those features.

Districts often communicated program information through principals, who could influence how information about the TIF program in their schools reached teachers. Therefore the survey asked both teachers and principals about teachers' eligibility for pay-for-performance bonuses as well as the magnitude of the bonus.

Educator reports revealed several general patterns. First, most teachers and principals reported that their schools implemented all required components except pay-for-performance—evaluations based on achievement growth and classroom observations, opportunities for additional pay, and professional development. However, educators often misunderstood the performance measures and the pay-for-performance bonuses used for TIF. Second, for the key component—pay-for-performance, which was expected to distinguish the treatment and control groups—educators in treatment schools reported, as expected, greater rates of eligibility than those in control schools. However, the size of the differences was smaller than intended under the TIF grant. Third, consistent with the grant design, educators in treatment and control schools reported similar understanding of most TIF components other than pay-for-performance. We discuss these findings in greater detail next.

Educator Versus District Reports of TIF Implementation

Educators' perceptions of several components of TIF besides pay-for-performance could shape the eventual impact of pay-for-performance on their behavior. In particular, educators' beliefs about the performance measures on which they were being evaluated could determine their understanding of how their compensation could be tied to performance. Their participation in professional development focusing on understanding these measures might influence their capacity to improve their performance ratings in response to any incentives. Finally, their perceptions of opportunities for additional pay other than pay-for-performance might indicate, more broadly, their understanding of the full compensation package that TIF offers. If educators were aware of these additional compensation opportunities, pay-for-performance could factor less in shaping their decisions. Below, we discuss educators' reports of their exposure to each of these components and compare them with districts' reports.

In evaluation districts, educators' reported awareness of evaluation measures often differed from districts' reports. Two-thirds (68 percent) of teachers reported being evaluated on achievement growth measures, and nearly four-fifths (78 percent) reported being evaluated through formal observations (Table IV.7). Nevertheless, these percentages were lower than suggested by district reports, which indicated that all evaluation districts used both types of measures to evaluate teachers in TIF schools. Because teachers in nontested grades and subjects could not be evaluated on achievement growth in their own classrooms, we examined teachers in tested and nontested grades and subjects separately, to determine whether they perceived being evaluated differently. Teachers in tested grades and subjects were more likely than those in nontested grades and subjects

³⁶ We present findings on the optional components in Appendix D, Tables D.7–D.9.

to report being evaluated on measures of student achievement growth (75 versus 64 percent; Appendix D, Table D.11). However, reports of classroom observations were similar in the two groups.

Principals' awareness of the measures used to evaluate teachers in their schools varied by the type of measure. Fifty-six percent of principals reported that achievement growth measures were used to evaluate teachers in their schools, significantly lower than the percentages reported by teachers (68 percent) and districts (100 percent). However, consistent with all evaluation districts implementing formal observations, nearly all principals (98 percent) were aware that those measures were used to evaluate teachers.

Overall, these findings indicate that districts were able to inform most teachers and principals about the measures on which teachers would be evaluated. For the remaining educators, several scenarios may explain the differences between their reports and district reports. One possibility is that district communications about those measures did not reach all of the educators in its intended audience. Although we do not have individual teacher data on the extent and coverage of districts' communication efforts, we asked teachers about their exposure to professional development focusing on the performance measures. As discussed in Appendix D, teachers' exposure to this type of support was lower than expected given district reports. This finding suggests that at least one mode of communication was not reaching all teachers.

Other possibilities, which we do not have the data to assess, are that information about the performance measures was not communicated clearly, or that educators did not pay attention to or did not understand this information. The data we do report are based on what was, in most cases, the first year of implementation of a new program, before educators had actually received information based on the effectiveness measure used for TIF. In future reports, once everyone involved in TIF gains implementation experience and educators receive more feedback on how they performed on the measures, we will be able to reassess the consistency between district and teacher reports.

In Table IV.7, we also show that teachers reported more formal classroom observations per year than principals and districts. On average, teachers reported that they would be observed nine times by the end of the 2011–2012 school year, compared to three times reported by principals and four times by districts. This difference is primarily driven by a small fraction of teachers—about 10 percent—who reported 20 or more formal observations. The median number of formal observations reported by teachers was 5.

Principals' reports of how they were evaluated were largely consistent with their districts' reports. Most principals (89 percent) reported that they were evaluated on the basis of student achievement growth for their entire school (Table IV.8). These reports deviated only slightly from all districts reporting using those measures to evaluate principals. For the remaining measures—student achievement levels and achievement growth in certain student groups—the percentages of principals and districts reporting that the measure was used were statistically indistinguishable from each other.

Most teachers reported receiving professional development about the TIF program as required by the TIF grant. Under the TIF grant requirements, an intended use of TIF funds was to support professional development that would help teachers to understand the TIF program, particularly the performance measures on which they were being evaluated. Most teachers reported receiving professional development on these topics. However, teachers' reported participation rates

in such activities were lower than expected, given their districts’ reports. In Appendix D (Table D.10), we present these findings.

Table IV.7. Performance Measures Used to Evaluate Teachers, as Reported by Educators and District Representatives (percentages unless otherwise noted)

Performance Measure	Percentages of Respondents Reporting the Measure Was Used		
	Teacher Report	Principal Report	District Report
Student Achievement			
Student achievement level	56.8 +	43.7	30.0
Student achievement growth	68.0*+	56.3*	100.0
By school	61.7*	51.1*	100.0
By student group	50.3 +	36.5	44.4
By teacher’s classroom	58.2	53.3	60.0
Classroom Observations	78.1*+	97.5	100.0
Number of classroom observations per year (Averages)	8.9*+	3.3	3.6
Sample Size—Range^a	648–822	112–134	9–10

Source: Teacher, principal, and district surveys.

Notes: Overall values for teacher and principal responses are weighted means so that districts are equally weighted. Overall values for districts are means among the 10 evaluation districts that participated in the educators’ survey. Educators’ responses are included only if their district responded to the given question. Classroom observations are standardized by using a rubric or checklist and are usually given at regular intervals.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference from the district report is statistically significant at the 0.05 level, two-tailed test.

+Difference between teacher and principal report is statistically significant at the 0.05 level, two-tailed test.

Table IV.8. Performance Measures Used to Evaluate Principals, as Reported by Principals and District Representatives

Performance Measure	Percentage of Respondents Reporting the Measure Was Used	
	Principal Report	District Report
Student achievement level	85.6	60.0
Student achievement growth for the school	88.7*	100.0
Student achievement growth in certain student groups	84.7	60.0
Sample Size—Range^a	123–128	10

Source: Principal and district surveys.

Note: Overall values for principal responses are weighted means so that districts are equally weighted. Overall values for districts are means among the 10 evaluation districts that participated in the educators’ survey. Educators’ responses are included only if their district responded to the given question.

^aSample 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.

Most teachers reported receiving professional development about the TIF program as required by the TIF grant. Under the TIF grant requirements, an intended use of TIF funds was to support professional development that would help teachers to understand the TIF program, particularly the performance measures on which they were being evaluated. Most teachers reported receiving professional development on these topics. However, teachers’ reported participation rates in such activities were lower than expected, given their districts’ reports. In Appendix D (Table D.10), we present these findings.

Teachers were less likely than their principals and district representatives to report that teachers could earn additional pay for extra roles and responsibilities. TIF districts were required to offer opportunities for educators to earn additional pay for taking on extra responsibilities. Most teachers (61 percent) reported that they or their colleagues in the same school were eligible to earn additional pay for extra responsibilities (Table IV.9). However, all evaluation districts reported offering this type of additional pay to teachers, and the discrepancy between teacher and district reports about this opportunity was statistically significant. All districts said they offered additional pay for a category of responsibilities known variously as mentor, master, or lead teacher, but only 56 percent of teachers reported that additional pay was available to teachers in their schools for this type of responsibility.

Principals were more informed about their teachers’ eligibility to earn compensation for roles and responsibilities than the teachers themselves were (86 percent of principals reported that these opportunities were available to teachers in their schools). Nevertheless, principals’ knowledge of these opportunities did not lead to similar levels of awareness among teachers.

Table IV.9. Additional Teacher Pay for Extra Roles and Responsibilities, as Reported by Educators and District Representatives

Role or Responsibility	Percentage of Respondents Reporting Teachers Can Receive Additional Pay for the Specified Role or Responsibility		
	Teacher Report	Principal Report	District Report
Any added roles or responsibility	61.4**	86.4*	100.0
Mentor, master, or lead teacher	55.5**	85.2*	100.0
Department chair or head	23.2	18.1	30.0
Lead curriculum specialist	24.2	26.4	11.1
Member of schoolwide committee or task force	13.0	7.6	20.0
Member of leadership team	32.3	24.6	22.2
Sample Size—Range^a	486–815	126–135	9–10

Source: Teacher, principal, and district surveys.

Notes: Overall values for teacher and principal responses are weighted means so that districts are equally weighted. Overall values for districts are means among the 10 evaluation districts that participated in the educators’ survey. Educators’ responses are included only if their district responded to the given question.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference from the district report is statistically significant at the 0.05 level, two-tailed test.

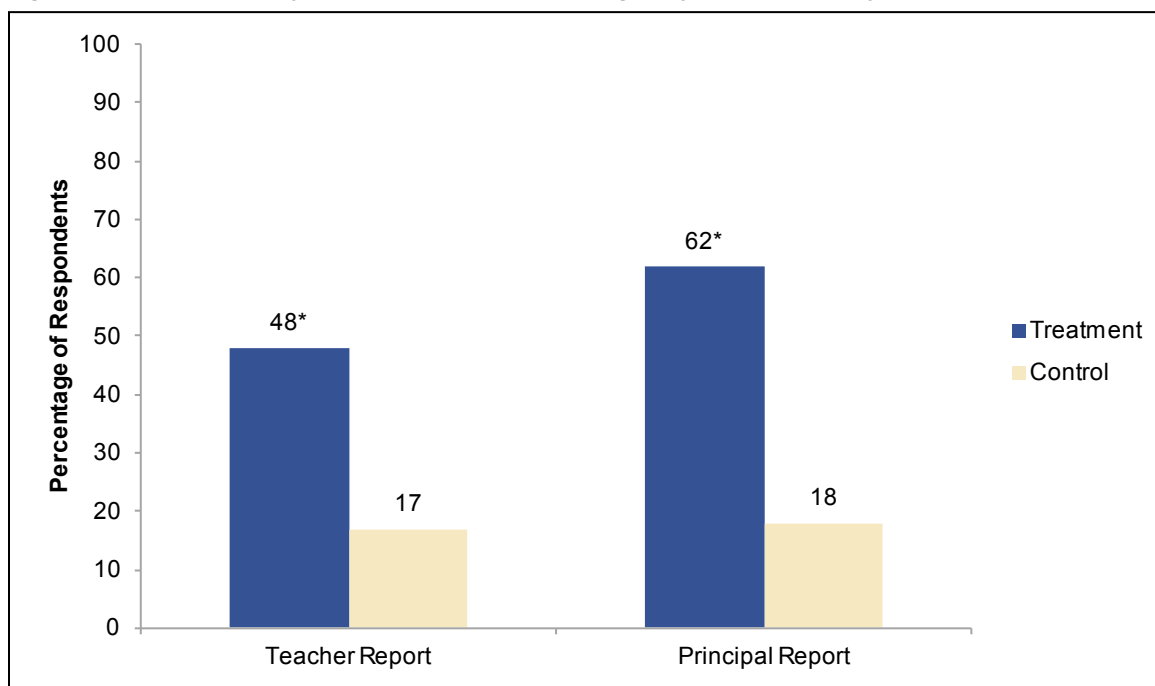
+Difference between teacher and principal report is statistically significant at the 0.05 level, two-tailed test.

Educators’ Understanding of Pay-for-Performance and Automatic Bonuses, in Treatment and Control Schools

According to the study design, educators in treatment schools were eligible for pay-for-performance bonuses and educators in control schools were eligible for automatic bonuses. Educators’ understanding about these bonuses is, therefore, critical to the program’s impact on educators.

Fewer than half of the teachers in treatment schools thought they were eligible for pay-for-performance bonuses. In both treatment and control schools, we found inconsistencies between eligibility for pay-for-performance and teachers’ reports of such eligibility. First, in treatment schools, fewer than half (48 percent) of teachers believed that they or their colleagues in the same school were eligible for pay-for-performance bonuses (Figure IV.7). Second, in control schools, although teachers were not supposed to be eligible for pay-for-performance bonuses, 17 percent of teachers reported that they or their same-school colleagues were eligible (Figure IV.7).³⁷

Figure IV.7. Teachers’ Pay-for-Performance Bonus Eligibility, as Reported by Teachers and Principals



Source: Teacher and principal surveys.

Note: Figures indicate the percentage of respondents who reported that teachers in their schools were eligible for pay-for-performance. A total of 395 treatment teachers, 392 control teachers, 67 treatment principals, and 65 control principals responded to this survey question.

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

³⁷ From technical assistance documents and telephone interviews with district staff, we found that three districts actually offered small pay-for-performance bonuses to all schools in those districts, including control schools. In those districts, 20 percent of teachers in control schools reported that they were eligible for pay-for-performance (Appendix D, Table D.12). However, even in the remaining districts, 17 percent of teachers in control schools believed they were eligible for pay-for-performance when it was not offered in their school. Therefore, the existence of districtwide pay-for-performance programs in the three districts does not account for most of the inconsistency between teachers’ reports of pay-for-performance eligibility in control schools and the schools’ assigned ineligibility for pay-for-performance.

We asked principals to report on the eligibility of teachers in their schools to receive pay-for-performance (see Figure IV.7 for results). Our study revealed a familiar pattern: unlike the assigned status, not all treatment school principals reported pay-for-performance eligibility (62 percent instead of the expected 100 percent), and some control school principals reported such eligibility (18 percent instead of the expected 0 percent).

Not surprisingly, these patterns of perception do not translate into a treatment-control difference in reported pay-for-performance eligibility of 100 percentage points. The differences we did observe were statistically significant: 31 percentage points, according to teachers, and 44 percentage points, according to principals (Figure IV.7). Nevertheless, these differences were much smaller than intended under the TIF grant.

Both teachers and principals underestimated the maximum pay-for-performance bonus for which teachers were eligible. Teachers' perceptions about the maximum size of the pay-for-performance bonuses could influence them to orient their behaviors toward obtaining one. To gauge the strength of the incentives perceived by teachers in treatment schools compared with those in control schools, we measured differences between these groups in the maximum pay-for-performance bonus for which they thought teachers in their school were eligible. We included all teachers in these analyses; for teachers who reported that they and their school colleagues were ineligible for pay-for-performance, we recorded their response as indicating a maximum pay-for-performance bonus of zero.

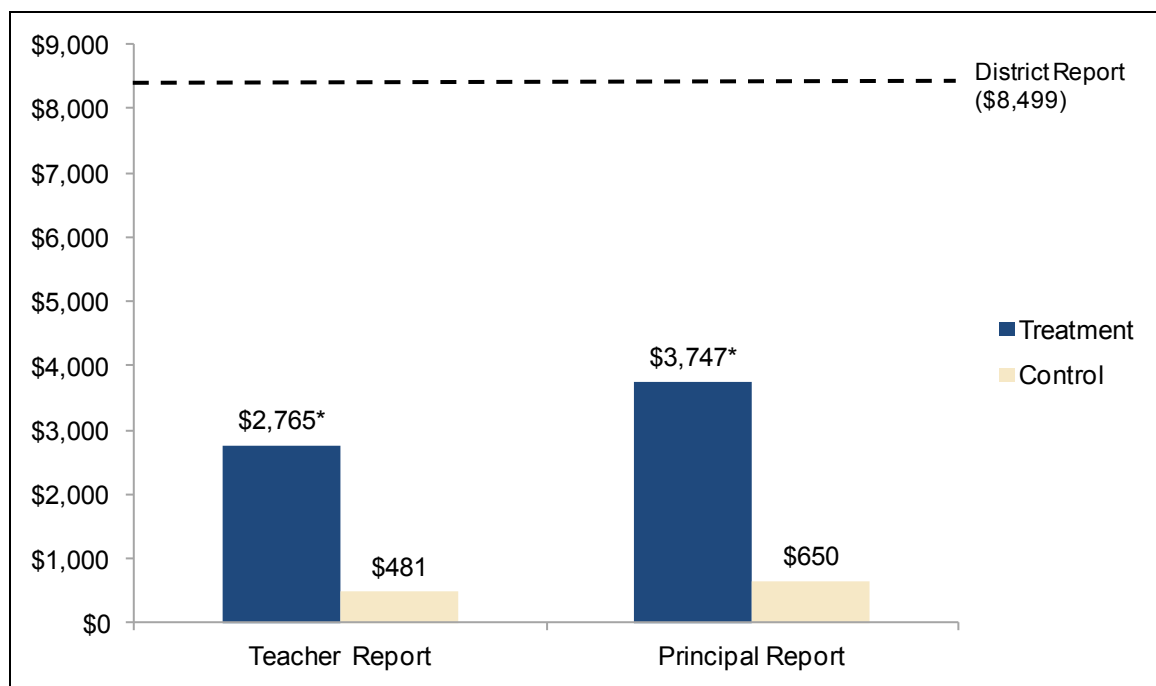
Teachers in treatment schools believed that they or their colleagues in the same school were eligible for a larger maximum pay-for-performance bonus than did teachers in control schools. On average, teachers in treatment schools perceived a maximum pay-for-performance bonus of about \$2,800; those in control schools perceived a maximum pay-for-performance bonus of about \$500, a statistically significant difference of \$2,300 (Figure IV.8).

Nevertheless, teachers in treatment schools underestimated the maximum pay-for-performance bonus amount for which, according to district survey responses, they or their school colleagues were eligible. As described earlier in this chapter, evaluation districts expected to offer, on average, a maximum pay-for-performance bonus amount of \$8,499—more than three times the amount perceived by teachers in treatment schools. One reason, as we show in Figure IV.7, is that fewer than half of these teachers thought they and their colleagues were eligible for *any* pay-for-performance bonus. Yet, even teachers in treatment schools who thought that they *were* eligible for pay-for-performance bonuses underestimated the maximum size of those bonuses; they believed that the maximum amount was about \$5,800.³⁸

Principals in treatment schools also underestimated the maximum pay-for-performance bonus amount for which their teachers were eligible. On average, principals in treatment schools believed that their teachers could earn just over \$3,700 from pay-for-performance—less than half of the amount reported by districts (Figure IV.8).

³⁸ This amount was calculated as the ratio of two numbers: (1) the average maximum pay-for-performance bonus amount perceived by all teachers in treatment schools (\$2,765; see Figure IV.8), and (2) the fraction of teachers in treatment schools who understood that they or their same-school colleagues were eligible for pay-for-performance bonuses (0.48; see Figure IV.7).

Figure IV.8. Maximum Possible Size of Pay-for-Performance Bonuses for Teachers, as Reported by Teachers and Principals



Source: Teacher and principal surveys.

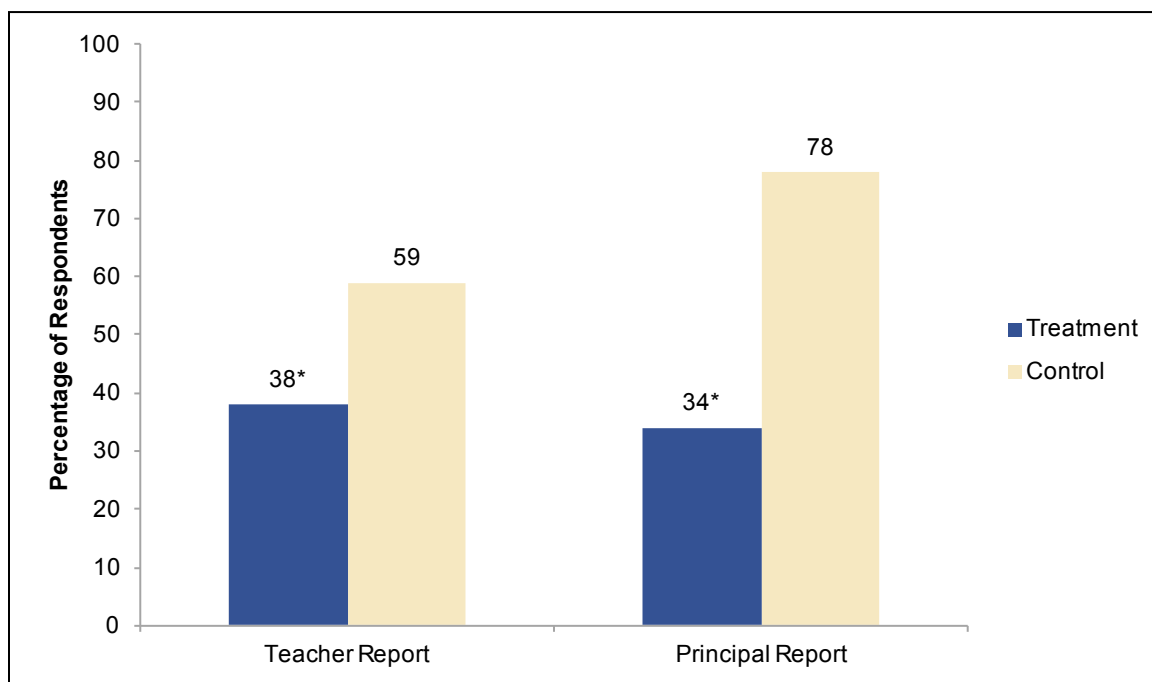
Notes: Figures indicate respondents' average report of the maximum possible size of teachers' pay-for-performance bonuses. A total of 348 treatment teachers, 378 control teachers, 53 treatment principals, and 58 control principals responded to this survey question. The maximum bonus amount was set to zero for all respondents who indicated they and their school colleagues were ineligible for a bonus. For educators who reported being eligible for the bonus but did not indicate an amount, bonus amounts were imputed through multiple imputation methods. This approach led to 47 additional responses for treatment teachers, 14 for control teachers, 14 for treatment principals, and 7 for control principals, bringing the total sample size to 395 treatment teachers, 392 control teachers, 67 treatment principals, and 65 control principals. See Appendix C for additional discussion on the imputation method. Figure D.2 in Appendix D shows that our results are similar if we do not impute the missing bonus amounts.

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

Teachers' reports of eligibility for automatic bonuses differed between treatment and control schools, but not by as much as intended under the study design. Evaluation districts were supposed to give teachers in control schools a small, automatic bonus equal to approximately 1 percent of the teacher's salary for their school's participation in TIF. Because the average self-reported base salary for teachers in our study was about \$47,000, an average automatic bonus consistent with the TIF grant would have been approximately \$500. As with pay-for-performance bonuses, teachers' perceptions about their eligibility for automatic bonuses indicate their awareness of the compensation that was unique to their school's treatment status.

Teachers in control schools were more likely than teachers in treatment schools to believe that they or their same-school colleagues were eligible for automatic bonuses (59 versus 38 percent; Figure IV.9). Nevertheless, this result is not the 100 percentage point difference you might have expected from the study design. In both treatment and control schools, about two-fifths of the teachers’ understanding about their eligibility for automatic bonuses did not match their schools’ treatment status. These inconsistencies occurred despite the fact that the automatic bonus had a simple, across-the-board structure. Therefore, it is possible that the greater complexity of the pay-for-performance bonus was not the primary source of the inconsistencies between teachers’ perceptions of pay-for-performance eligibility and their schools’ treatment status. For both types of bonuses, the data suggest that many teachers did not receive, pay attention to, or understand basic information about how their school’s treatment status affected their compensation.

Figure IV.9. Teachers’ Automatic Bonus Eligibility, as Reported by Teachers and Principals



Source: Teacher and principal surveys.

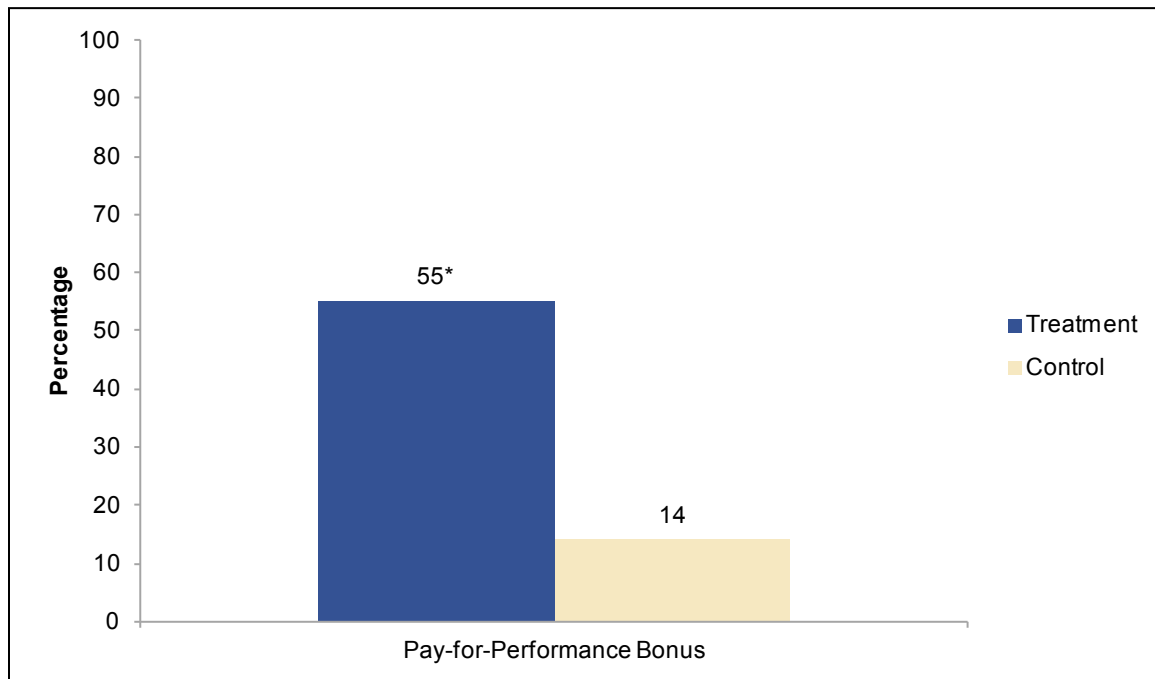
Note: Figures indicate the percentage of respondents who reported that teachers in their schools were eligible for automatic bonuses. A total of 396 treatment teachers, 394 control teachers, 67 treatment principals, and 65 control principals responded to this survey question.

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

Most principals’ reports about their own eligibility for pay-for-performance and automatic bonuses were consistent with their schools’ treatment status, but the difference was smaller than intended under the TIF grant. For pay-for-performance bonuses, 55 percent of principals in treatment schools reported that they were eligible, compared with 14 percent of principals in control schools (Figure IV.10). The resulting treatment-control difference, 41 percentage points, was less than half of the intended difference of 100 percentage points. Conversely, the percentage of principals who believed they were eligible for automatic bonuses was higher in control schools than in treatment schools—by 39 percentage points (Figure IV.11).

Like teachers, principals underestimated the maximum amount of the pay-for-performance bonus for which district representatives indicated they were eligible. On average, principals of treatment schools thought that they could earn up to about \$4,700 in pay-for-performance bonuses (Figure IV.12). In contrast, as shown earlier in this chapter (see Figure IV.2), evaluation districts reported that the average maximum expected pay-for-performance bonus for principals was \$9,600—more than twice the maximum amount that principals of treatment schools had estimated.

Figure IV.10. Principals’ Reports of Their Own Eligibility for Pay-for-Performance Bonuses

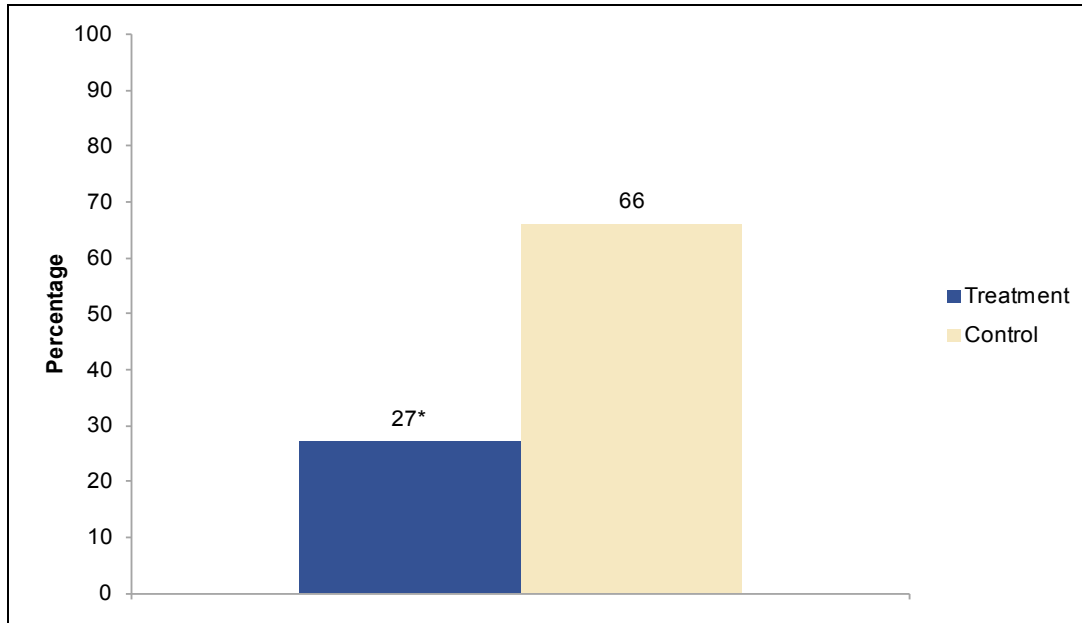


Source: Principal survey.

Note: Figures indicate the percentage of principals who reported that they were eligible for pay-for-performance. A total of 67 treatment principals and 66 control principals responded to the survey question.

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

Figure IV.11. Principals' Reports of Their Own Eligibility for Automatic Bonuses

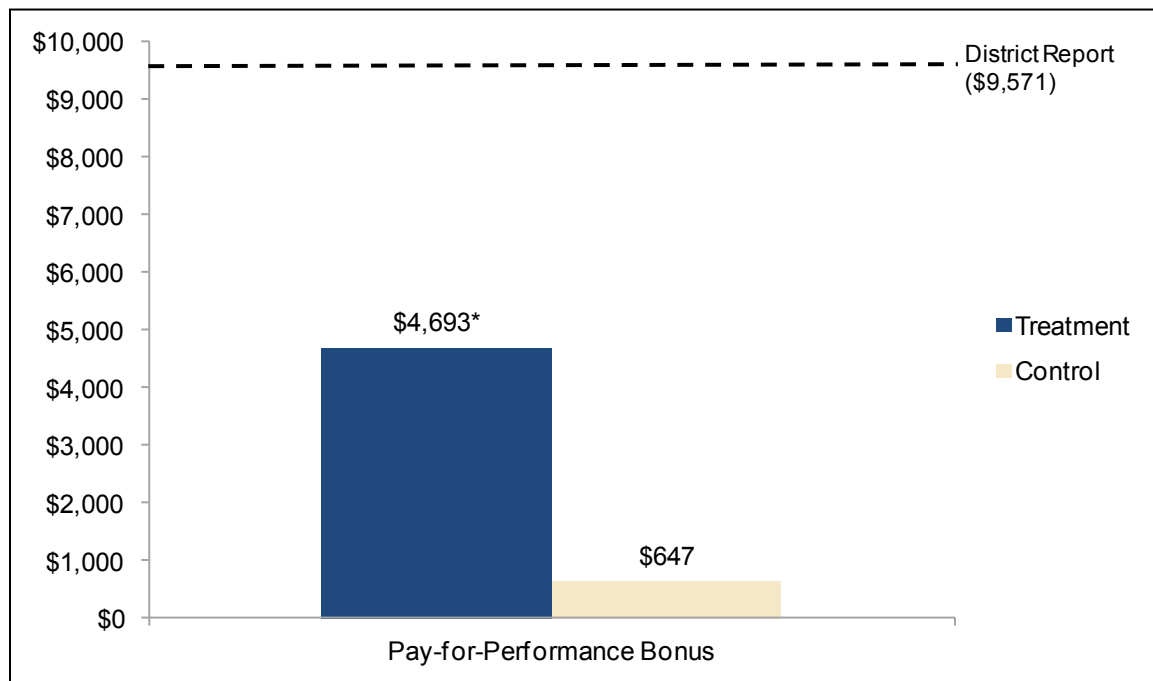


Source: Principal survey.

Note: Figures indicate the percentage of principals who reported that they were eligible for automatic bonuses. A total of 66 treatment principals and 67 control principals responded to the survey question.

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

Figure IV.12. Principals' Reports of the Maximum Possible Size of Their Pay-for-Performance Bonuses



Source: Principal survey.

Notes: Figures indicate principals' average report of the maximum possible size of their pay-for-performance bonuses. Fifty-nine treatment principals and 62 control principals responded to the survey question on pay-for-performance bonuses. The maximum bonus amount was set to zero for all principals who indicated they were ineligible for a bonus. For principals who reported being eligible for the bonus but did not specify a bonus amount, bonus amounts were imputed through multiple imputation methods. This approach led to eight additional responses from treatment principals and four from control principals, bringing the total sample size to 67 treatment principals and 66 control principals. See Appendix C for additional discussion. Figure D.3 in Appendix D shows that our results are similar if we do not impute the missing bonus amounts.

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

Implementation of Other TIF Components, by Treatment Status

Although eligibility for pay-for-performance and automatic bonuses was the only component of TIF that was supposed to differ between treatment and control schools, we also compared schools' implementation of other program components, as reported by teachers and principals. Evaluation districts might have had reason to implement these other components differently in treatment and control schools, despite the intentions of the study design. For example, districts could have expanded eligibility for additional pay in control schools more than in treatment schools as a way of compensating control schools for their ineligibility to earn pay-for-performance bonuses. Moreover, performance measures were more consequential in treatment schools than in control schools. Their use as inputs into pay-for-performance bonuses could have led to different performance evaluations in the two groups.

For most of the performance measures that we examined in the study, similar percentages of educators in treatment and control schools reported being evaluated on the measure. For example, slightly more than three-fifths of teachers in both groups reported being evaluated on student achievement growth for the entire school; nearly 9 of 10 teachers in both groups reported being evaluated based on principals' or other administrators' professional judgment (Figure IV.13). For measures based on student achievement levels and those based on classroom observations, control teachers were more likely than treatment teachers to report being evaluated on the measures, but these differences did not exceed 6 percentage points. On every measure of principal performance, treatment and control schools were statistically indistinguishable in the percentage of principals who reported being evaluated on the measure (Figure IV.14).

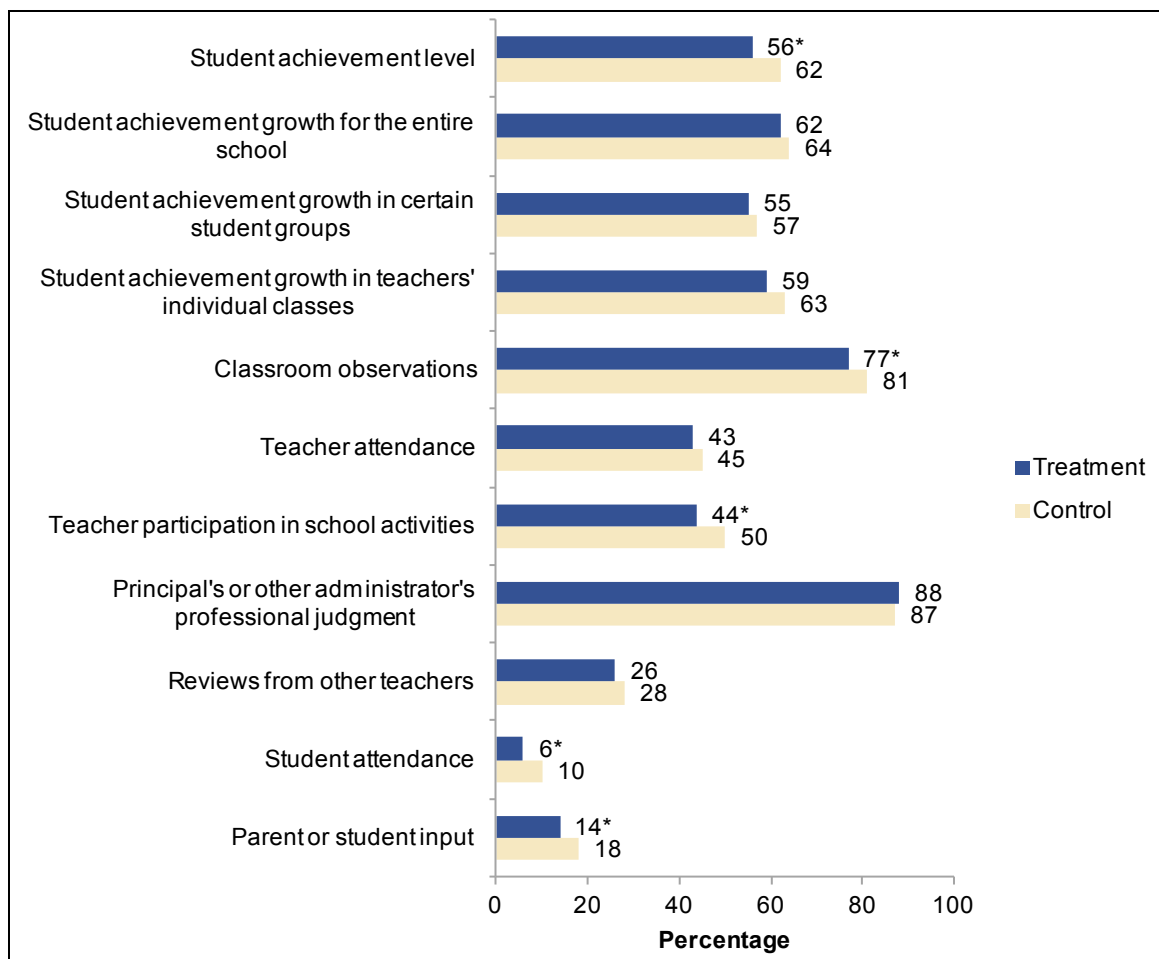
Teachers in treatment and control schools reported similar rates of eligibility for additional pay as intended by the study design. On nearly every opportunity for additional pay other than pay-for-performance and automatic bonuses, similar percentages of teachers in treatment and control schools reported being eligible (Figure IV.15). Nearly three-fifths of teachers in both groups reported opportunities to earn additional pay by taking on extra responsibility. Slightly over one-fifth of teachers reported that they or their colleagues in the same schools could earn additional pay by teaching in a hard-to-staff school, but this was not a requirement of the grant.

Teachers in treatment schools reported participating in more hours of professional development during the first half of the school year than teachers in control schools. Teachers were asked to report their participation in professional development from July 1 to December 31 on a range of topics, including understanding components of TIF as well as more general topics such as differentiating instruction and aligning curricula to standards. Teachers in treatment and control schools reported similar rates of participation in professional development activities focused on understanding components of TIF (Table IV.10). We found some treatment-control differences in participation rates for professional development focused on more general topics, but the direction of those differences was inconsistent. A more consistent pattern emerged, however, in the numbers of hours teachers reported they spent on professional development. Teachers in treatment schools reported spending more hours of professional development on all of the topics than teachers in control schools. In total, teachers in treatment schools reported spending seven hours more on professional development activities during the first half of the school year than their counterparts in control schools (Table IV.11).

The extent of teachers' participation in professional development could reflect both the activities offered or required by their districts and the amount of optional professional development in which teachers chose to enroll. Therefore, the treatment-control differences in reported hours of

professional development, shown in Table IV.11, could be due to either differences in how districts implemented professional development in treatment and control schools or teachers' own responses to pay-for-performance. Notably, teachers' reports of professional development from July 1 to December 31 presumably included any professional development mandated by district policy before the start of the school year. Therefore, responses about professional development later in the school year may better capture teachers' individual choices to pursue professional development in response to eligibility for pay-for-performance. In Chapter V, we report the impact of pay-for-performance on the amount of time teachers had spent in professional development at the time of the spring survey.

Figure IV.13. Teachers' Reports of the Measures Used to Evaluate Their Own Performance

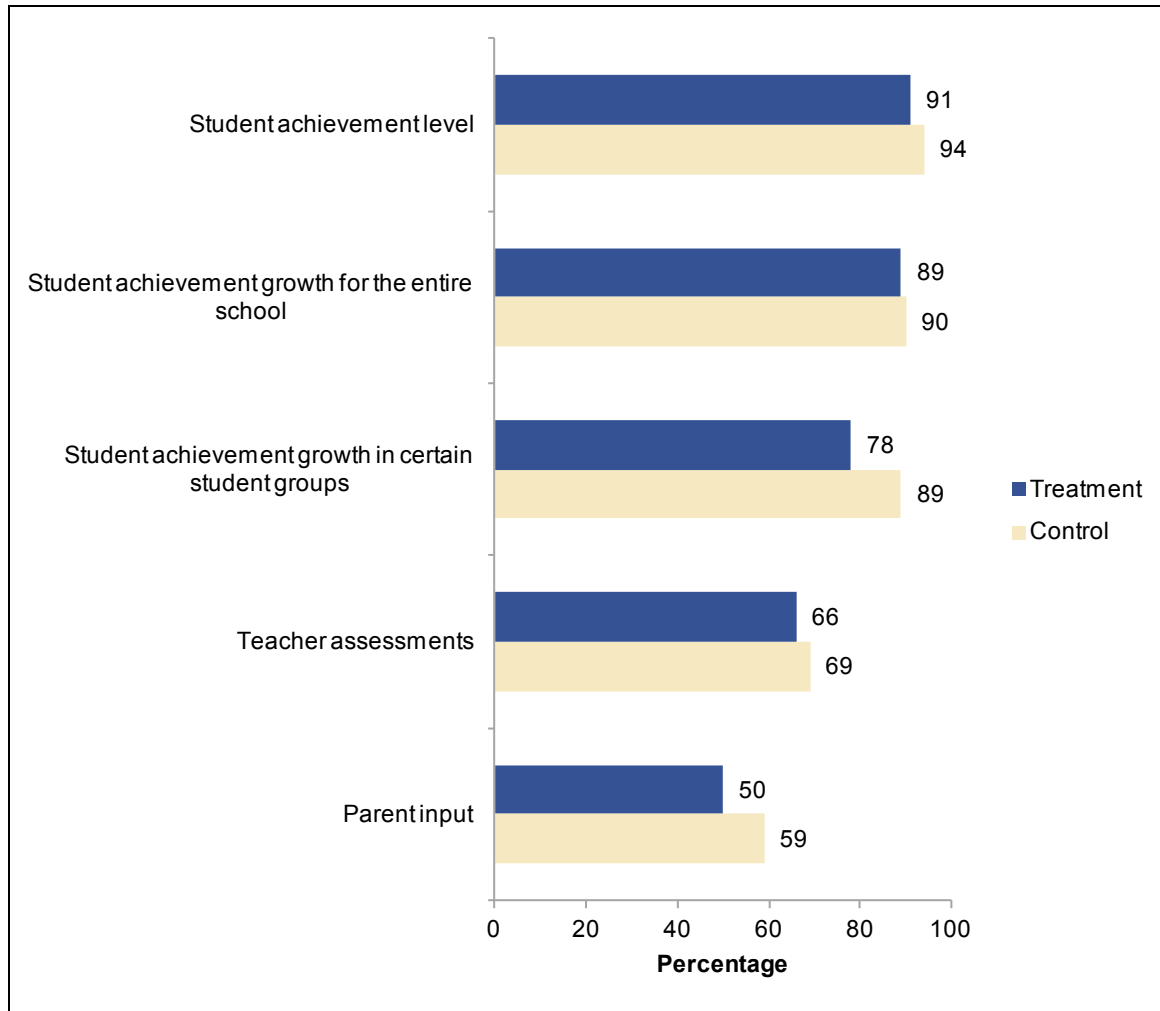


Source: Teacher survey.

Note: Figures indicate the percentage of teachers who reported that the specified measure was used to evaluate their performance. A total of 400 treatment teachers and 408 control teachers responded to this survey question.

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

Figure IV.14. Principals' Reports of the Measures Used to Evaluate Their Own Performance

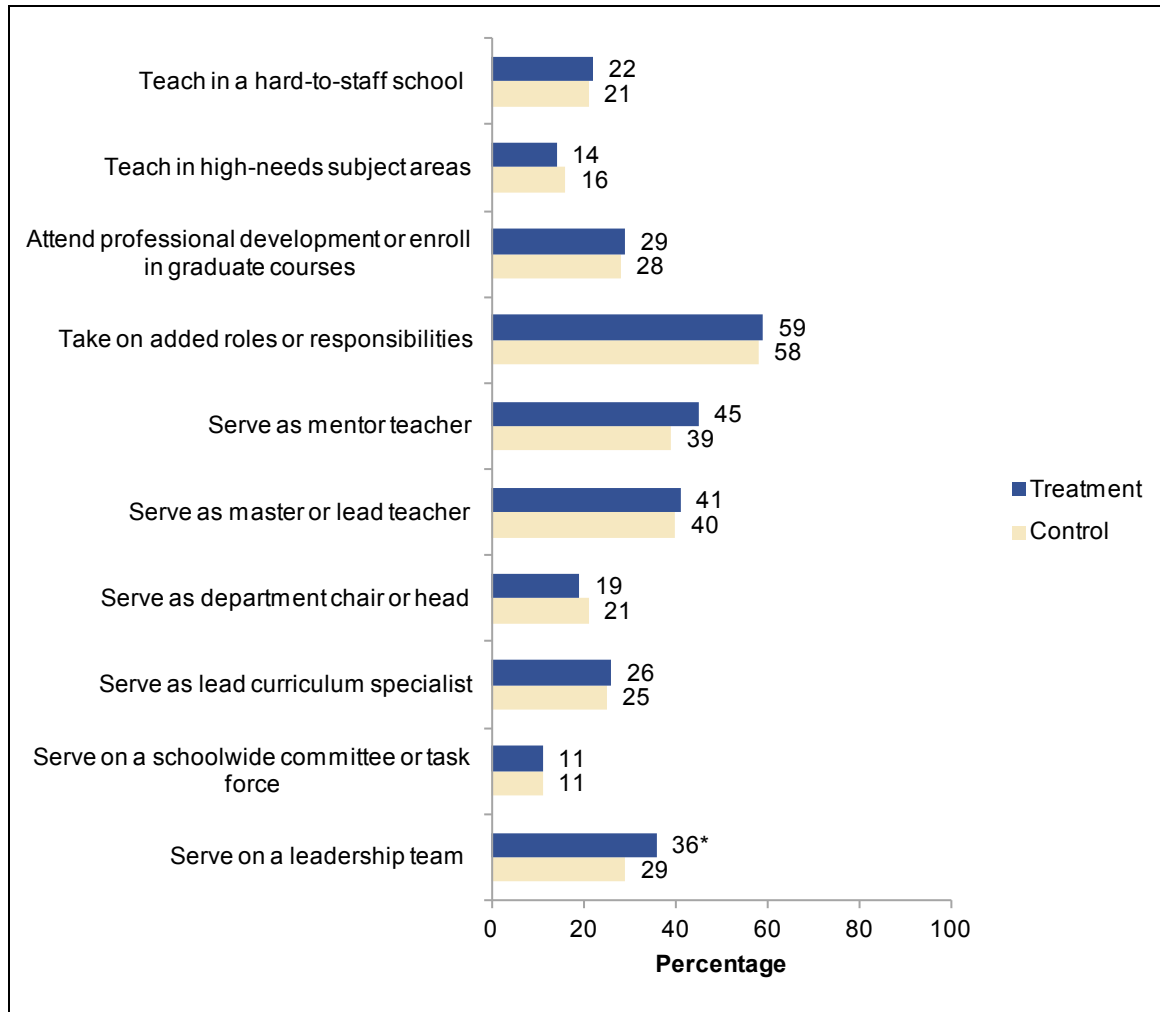


Source: Principal survey.

Note: Figures indicate the percentage of principals who reported that the specified measure was used to evaluate their performance. A total of 55 treatment principals and 58 control principals responded to this survey question.

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

Figure IV.15. Teachers' Reports of Whether Teachers in Their Schools Were Eligible for Additional Pay Opportunities



Source: Teacher survey.

Note: Figures indicate the percentage of teachers who reported that teachers in their school were eligible for additional pay for the specified reason. A total of 253 treatment teachers and 243 control teachers responded to this survey question.

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

Table IV.10. Teachers' Reports of the Professional Development They Received (percentages)

Topic of Professional Development	Overall	Treatment	Control	Treatment-Control Difference
Understanding components of TIF program	77.3	79.2	75.6	3.6
Understanding performance measures of TIF program	73.6	74.5	73.0	1.5
Feedback based on TIF performance ratings	53.8	52.4	55.0	-2.7
Differentiated instructional strategies based on student assessments	70.7	68.0	73.5	-5.6*
Instructional techniques and strategies	85.0	87.1	83.2	3.9*
Aligning curricula to state or district standards	79.8	78.7	81.2	-2.5
Number of Teachers—Range^a	802–805	401–402	400–404	

Source: Teacher survey.

Note: Figures indicate the percentage of teachers who reported receiving professional development in the specific topic between July 1, 2011, and December 31, 2011.

^aSample 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 IV.11. Teachers' Reports of Hours Spent in Professional Development Activities (averages)

Topic of Professional Development	Overall	Treatment	Control	Difference
Understanding components of TIF program	5.2	6.3	4.0	2.3*
Understanding performance measures of TIF program	3.4	3.8	2.9	0.9*
Feedback based on TIF performance ratings	2.2	2.4	2.0	0.3
Differentiated instructional strategies based on student assessments	6.0	6.2	5.8	0.5
Instructional techniques and strategies	11.1	12.2	10.0	2.1*
Aligning curricula to state or district standards	8.7	9.4	8.0	1.5*
Total hours on professional development	46.1	50.0	42.5	7.4*
Number of Teachers—Range^a	770–812	382–406	387–406	

Source: Teacher survey.

Note: Figures indicate teachers' average reported hours spent on professional development in the specific topic between July 1, 2011, and December 31, 2011.

^aSample 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

V. INTERMEDIATE IMPACTS ON EDUCATORS' ATTITUDES AND BEHAVIORS

The final aspect of pay-for-performance bonuses we examined for this report was the effect on educators' attitudes, such as job satisfaction, and behaviors, such as how teachers allocate time during the school day and whether they plan to remain at their schools. As shown in the theory of change in Chapter I, 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' attitudes and behaviors, both of which may affect educators' productivity and school choice. Because both treatment and control schools offered all

required components of the Teacher Incentive Fund (TIF) program except pay-for-performance, we can estimate the impact of pay-for-performance by comparing the responses of educators in treatment schools to those in control schools. These findings are based on the first year of implementation, when the pay-for-performance program was new, and educators may have not yet received all the information on their performance on the effectiveness measures such as student achievement growth, and the first round of bonus payments had not been made.

Impacts 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 toward the TIF program.

Satisfaction

Most teachers in both treatment and control schools were satisfied with professional opportunities, school environment, and their jobs overall. Table V.1 presents the percentages of teachers in treatment and control schools who were "somewhat satisfied" or "very satisfied" with several aspects of their jobs.

Key Findings on Early Impacts of Pay-for Performance

- **Most teachers and principals in treatment and control schools reported being satisfied with their professional opportunities and school environment.**
- **A lower percentage of teachers in treatment schools than in control schools reported that they were satisfied with professional opportunities, school environment, and the TIF program, but a higher percentage were satisfied with their opportunities to earn extra pay.**
- **A lower percentage of principals in treatment schools than in control schools reported that they were satisfied with school morale and with colleagues' contribution to student learning; yet principals' attitudes toward the TIF program were similar.**
- **Teachers in treatment schools reported spending more time on instruction than teachers in control schools, but not more time overall on other activities during school hours.**
- **Principals in treatment schools reported that TIF changed the way they recruited teachers to their schools, but not how they assigned staff in their schools.**
- **Teachers in treatment schools were more likely to report that TIF influenced their choice of where to teach, but only a small percentage of teachers or principals overall reported that TIF influenced their choices.**

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

Satisfaction Dimension	Treatment	Control	Impact
Use of Measures of Performance			
Classroom observations	68.4	77.0	-8.6*
Student achievement	65.4	67.4	-2.0
Opportunities for Pay and Development			
Opportunities for professional advancement	67.8	75.7	-7.8*
Opportunities to enhance skills	77.7	79.0	-1.3
Opportunities to earn extra pay	64.0	58.9	5.1*
School Environment			
Recognition of accomplishments	54.1	59.6	-5.4
Quality of interaction with colleagues	73.6	80.6	-7.0*
Colleagues' efforts	82.3	83.9	-1.6
School morale	48.1	54.9	-6.8*
Job Satisfaction			
Overall job satisfaction	67.6	72.9	-5.3
Number of Teachers—Range^a	405–409	405–412	

Source: Teacher survey.

^aSample 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.

A lower percentage of teachers in treatment schools than in control schools were satisfied with performance measures, professional opportunities, and school environment.

As shown in Table V.1, a smaller percentage of teachers in treatment schools than control schools were satisfied, on average, with the use of classroom observations as an evaluation measure (68 versus 77 percent), their opportunities for professional advancement (68 versus 76 percent), quality of interaction with colleagues (74 versus 81 percent), and school morale (48 versus 55 percent), and these differences were statistically significant. For most other satisfaction measures, the differences were also negative but not large enough to be statistically significant. The overall pattern of lower satisfaction among treatment teachers than control teachers had one exception: treatment teachers were more satisfied with opportunities to earn extra pay (64 versus 59 percent).³⁹

The bonuses could affect some groups of teachers differently, so we examined impacts separately by subgroup. Two subgroups 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, either 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, in some districts, teachers in tested grades and subjects were also able to earn greater maximum

³⁹ When we focused only on the percentage of teachers who were "very satisfied," a measure that is more sensitive to intensity of feeling, we found a similar pattern. The results are shown in Appendix E, Table E.1. We also present the results for educator satisfaction and attitudes toward TIF using different assumptions about weighting the effects across districts as well as a logit specification (Appendix E, Tables E.6 and E.7). The pattern of effects is similar under different modeling approaches, although they are no longer statistically significant under the alternative specifications.

pay-for-performance bonuses. In terms of experience, teachers who have been teaching longer under a different evaluation and compensation system may be less receptive to the new system.

We also examined subgroups of districts organized according to the features of their TIF program that we described in Chapter IV. We grouped districts by their relative emphasis on using student achievement growth in individual teachers' classrooms to evaluate teachers. This difference could affect the impact of pay-for-performance on teachers' satisfaction because teachers are evaluated differently and may be more or less likely to earn an award under different program types. We also looked at subgroups of districts based on the size of the maximum expected bonus. Teachers may respond more positively to the TIF program if they are eligible for larger bonuses. On the other hand, for those who believe that teachers should be paid similarly (or based on tenure), pay-for-performance with large payouts or large payout differentials may lower satisfaction.

The results of the subgroup analyses should be interpreted with caution. 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. 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 might have also 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.

In Table V.2, we show the estimated impacts of pay-for-performance for these four subgroups for the five satisfaction outcomes on which teachers in treatment and control schools demonstrated statistically significant differences overall. We found statistically significant impacts for teachers in tested grades and subjects, among teachers at all experience levels, in districts of all program types, and in districts with high pay-for-performance bonus amounts, but not for all five satisfaction measures shown in the table. Differences in impacts between subgroups were generally not statistically significant, with two exceptions: (1) more experienced teachers in treatment schools reported much lower satisfaction on most measures, and (2) teachers in treatment schools in districts that used a combination of teacher and school achievement growth in their performance measures (the Teacher Advancement Program (TAP) districts) were more satisfied on most measures. Appendix E, Table E.12 presents details of these hypothesis tests along with findings for five other teacher-satisfaction measures (such as opportunities to advance one's skills and overall satisfaction) for which the full sample impacts were not statistically significant.

Most principals in treatment and control schools were satisfied with professional opportunities, feedback, and the school environment. Table V.3 shows that the percentage of principals satisfied with several aspects of professional opportunities and school environment ranges from 71 to 100.

A lower percentage of principals in treatment schools than in control schools were satisfied with some dimensions of the school environment. Principals in treatment schools reported significantly lower satisfaction with school morale than principals in control schools (71 versus 88 percent) and were less likely to be satisfied with colleagues' contributions to student learning (94 versus 100 percent). However, when we look at the more intense response of principals who indicated they were "very satisfied," there were no statistically significant negative differences, and principals in treatment schools were significantly more likely to be satisfied with the opportunities to earn extra pay (see Appendix Table E.1).

Table V.2. Difference Between Teachers in Treatment and Control Schools on Selected Teacher-Satisfaction Measures, by Subgroup (Percentage Points)

Subgroup	Impacts on Whether Teachers Were "Somewhat" or "Very" Satisfied With...					
	Use of Classroom Observation	Opportunities for Professional Advancement	Opportunities to Earn Extra Pay	Quality of Interaction with Colleagues	School Morale	Number of Teachers
All Teachers (Primary)	-8.6*	-7.8*	5.1*	-7.0*	-6.8*	810
Teaching Assignment						
Tested grades and subjects	-8.9*	-8.8*	2.4	-10.4*	-7.3	485
Nontested grades and subjects	-7.7	-6.3	8.9	-1.6	-5.5	325
Teacher Experience						
Less than 5 years	-3.7	-13.8*	7.9	3.1	11.4	250
5 to 24 years	-8.8*	-5.7	6.5	-6.5*	-10.7*	482
More than 24 years	-20.4	-4.3	-13.7	-38.3*	-36.0*	77
District Program Type ^a						
No achievement growth measures by teacher	-13.3*	-7.7*	4.2	-10.3*	-17.1*	314
Emphasize achievement growth by teacher	-7.3	-11.5*	0.4	-5.4*	-4.8	374
Combine teacher and school growth (TAP) ^b	2.9	2.4	22.0*	0.6	20.4*	121
District Maximum Pay-for-Performance Bonus Amount ^c						
High (above median)	-11.6*	-9.6*	3.0	-10.4*	-9.3*	543
Low (below median)	2.9	-4.5	9.1	-0.3	-2.0	267

Source: Teacher survey, district survey, technical assistance (TA) documents, and district interviews.

^aProgram type classification was based on TA documents.

^bTAP = Teacher Advancement Program.

^cPay-for-performance bonus amount is calculated based on a combination of survey questions and district interviews, as described in Appendix C.

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

Table V.3. Principal Satisfaction with Professional Opportunities and School Environment (Percentages Who Were "Somewhat" or "Very" Satisfied)

Satisfaction Dimension	Treatment	Control	Impact
Opportunities for Pay and Development			
Opportunities to enhance skills	93.0	95.2	-2.2
Opportunities to earn extra pay	72.9	67.7	5.1
Intellectual challenge	97.3	96.8	0.5
Feedback on Performance	83.1	87.3	-4.2
School Environment			
Recognition of accomplishments	77.7	82.5	-4.8
Quality of interaction with colleagues	89.9	96.8	-6.9
Colleagues' efforts	92.6	98.4	-5.8
Colleagues' contribution to student learning	93.8	100.0	-6.2*
School morale	71.1	87.5	-16.4*
Number of Principals—Range^a	65–66	62–64	

Source: Principal survey.

^aSample 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.

Attitudes Toward TIF

A majority of teachers in both treatment and control schools were glad they were participating in TIF and thought the program was fair. Approximately two-thirds of teachers were glad they were participating in TIF; at least half believed TIF was fair (Table V.4). However, a lower percentage of teachers in treatment schools than in control schools believed that TIF was fair (53 versus 58 percent). Pay-for-performance eligibility increased the likelihood that teachers felt increased pressure to perform (63 versus 54 percent), and a lower percentage of teachers in treatment schools felt their principals were good judges of teacher talent (67 versus 74 percent). In addition, a lower percentage of treatment teachers than control teachers responded that TIF increased their job satisfaction (27 versus 32 percent).⁴⁰

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

Statement	Treatment	Control	Impact
Teachers who do the same job should receive the same pay	57.6	58.1	-0.5
Standardized student test scores in my district measure what students have learned	34.7	33.6	1.1
My principal is a good judge of teacher talent	66.5	73.6	-7.1*
I am glad that I am participating in the TIF program	67.0	64.9	2.1
My job satisfaction has increased due to the TIF program	27.1	32.0	-4.9*
I feel increased pressure to perform due to the TIF program	62.9	54.1	8.7*
I have less freedom to teach the way I would like to teach due to the TIF program	35.2	34.1	1.1
The TIF program has harmed the collaborative nature of teaching	24.8	23.5	1.3
The TIF program has caused teachers to work more effectively	48.2	44.4	3.8
The TIF program is fair	53.0	57.6	-4.6*
The process used to determine bonuses was adequately explained to me	67.8	60.1	7.8*
Number of Teachers—Range^a	399–406	394–410	

Source: Teacher survey.

^aSample 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.

In general, teachers in treatment and control schools had similar attitudes toward other aspects of their job and the TIF program, with one exception. Treatment teachers were more likely than control teachers to think that the process used to determine bonuses was adequately explained to them (68 versus 60 percent). On the other hand, pay-for-performance had no impact on teachers' perceptions about freedom to teach, the collaborative nature of teaching, or whether TIF caused teachers to work more effectively.

⁴⁰ When we focused only on treatment and control teachers who "strongly agreed," the percentages who responded that TIF increased their job satisfaction did not significantly differ. The results are shown in Appendix E, Table E.2.

Similar to the findings for satisfaction, we examined the impacts of pay-for-performance separately by subgroup for the five attitude outcomes on which teachers in treatment and control schools showed statistically significant differences (see Appendix E, Table E.13). Although most of the differences in impacts between subgroups were not statistically significant, there were some significant differences by teacher experience and program type. Novice teachers in treatment schools were more likely to report that the process used to determine bonuses was adequately explained, and more experienced teachers were less likely to report that TIF had increased their job satisfaction. Treatment teachers in districts that combined student achievement growth at the teacher and school levels (TAP districts) were more likely to report that the process used to determine bonuses was adequately explained. In districts that did not use student achievement growth in teachers' classrooms as a performance measure, treatment teachers were less likely to report that TIF increased their job satisfaction, that TIF was fair, and that their principals were good judges of teacher talent.

Principals in treatment and control schools had similar attitudes toward TIF. Unlike teachers, eligibility for pay-for-performance did not have any significant impacts on principals' attitudes toward TIF. As shown in Table V.5, we asked principals about their attitude toward several statements, such as whether the TIF program (1) contributed to greater teacher collaboration and (2) had been clearly communicated to them. We also asked principals whether they agreed with statements about broader TIF-related issues, such as whether TIF was likely to continue and whether they played an important role in implementing it in their school. The differences in responses between treatment and control principals were not statistically significant.⁴¹ More than 80 percent of principals in both groups reported that the TIF program was clearly communicated to them, that they played an important role in implementing it, and that the program was likely to continue in the future.

Table V.5. Principals' Attitudes Toward TIF Program (Percentages Who "Agreed" or "Strongly Agreed")

Statement	Treatment	Control	Impact
The TIF program has been clearly communicated to me	82.6	89.6	-7.0
This school has less chance of earning a bonus because of the characteristics of our student population	23.7	18.8	4.9
The evaluation system omits important aspects of school administration that should be considered	31.3	30.2	1.2
The TIF program contributes to greater collegiality and professionalism among the staff at this school	49.9	56.9	-7.0
Teachers at this school are more comfortable with frequent formal observations of their teaching because of the TIF program	56.1	61.5	-5.4
Parents and the school community believe the TIF program is important	39.6	46.9	-7.2
The TIF program is likely to continue for the foreseeable future	82.6	87.5	-4.9
I played an important role in implementing the TIF program at my school	82.0	83.1	-1.1
Number of Principals—Range^a	65–68	63–67	

Source: Principal survey.

^aSample sizes are presented as a range, based on the data available for each row in the table.

⁴¹ When we focused only on the percentage of principals who "strongly agreed," we found a similar pattern. The results are shown in Appendix E, Table E.3.

Impacts on Educators' Self-Reported Behaviors

In this section, we present estimates of the impact of pay-for-performance on educators' self-reported behaviors. We discuss impacts on two factors shown in the logic model in Chapter I: (1) educators' productivity and (2) recruitment and retention. Specifically, we examined how teachers allocated their time during the day, principals' recruitment of teachers, and teachers' and principals' decisions to move into or stay in a particular school. We do not yet have adequate data to measure the effectiveness of teachers who were recruited and retained, so we focus here on whether pay-for-performance affected teachers' mobility on average.

Teachers' Time Use

In spring 2012, we asked teachers to report how they spent their time in the most recent full week of teaching.

By the end of the first year of TIF implementation, teachers in treatment schools reported spending more time on instruction than teachers in control schools, but not more time during school hours overall. Because treatment school teachers were eligible to receive pay-for-performance bonuses that depended on student achievement growth, we hypothesized that they might use their time differently than those in the control group. For example, if test scores were used to determine bonuses, teachers might spend more time preparing their students for state tests. We found evidence suggesting that treatment school teachers spent more time on selected in-school activities. Treatment school teachers reported that they spent nearly 0.8 hours (48 minutes) more on classroom instruction in the most recent full week of teaching than control school teachers (Table V.6). However, the difference in the sum of time spent in all activities, including supervising students, prep time, and professional development, was not statistically significant.

Table V.6. Teachers' Average Time Spent on School-Related Activities in the Most Recent Full Week (Averages, in Hours)

	Treatment	Control	Impact
Time Spent During School Hours on			
Classroom instruction	27.1	26.3	0.8*
Supervising students in other activities	3.8	3.6	0.2
Class preparation and professional development with colleagues	9.7	10.0	-0.3
Other activities	2.3	1.6	0.7*
Hours absent	0.8	1.1	-0.3
Total time during school hours ^a (calculated)	43.1	41.5	1.6
Time Spent During Nonschool Hours on			
Academic activities with students	2.1	2.5	-0.3
Class preparation and professional development with colleagues	11.0	11.2	-0.2
Other school-related activities	1.7	2.0	-0.3
Total time during nonschool hours (calculated)	14.7	15.5	-0.8
Number of Teachers—Range^b	350–392	341–392	

Source: Teacher survey.

^aTotal time spent during school does not include self-reported time absent.

^bSample 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.

The findings here of no statistically significant differences in time spent on class preparation and professional development with colleagues contrast with the findings on time spent on professional development from Chapter IV (Tables IV.10 and IV.11). In Chapter IV, we found that teachers in treatment schools reported participating in more hours of professional development during the first half of the school year than teachers in control schools. Recall, the chapter IV findings are based on teachers' responses to professional development offered to them before—and during the first half of—the school year, whereas the results in this chapter are based on teachers' reported time use during the preceding week in the spring. Thus, the differences between teachers in treatment and control schools may be due to the structure and timing of the questions.

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 substantially higher bonuses in their schools than in control schools. In theory, being able to offer large bonuses might help principals recruit more teachers and higher-performing teachers.

Principals in treatment schools were more likely than principals in control schools to report using bonuses and the TIF program as a recruitment incentive. Although treatment and control principals emphasized similar points to recruit teachers (Table V.7), more treatment than control school principals reported using pay-for-performance to recruit teachers (26 versus 17 percent), and more treatment school principals reported using the TIF program as a recruitment incentive (46 versus 29 percent).⁴² These results reflect incentives that principals “often” or “always” used. Among incentives that were “always” used, treatment principals were also significantly more likely to report using opportunities for career advancement to recruit teachers (Appendix E, Table E.4).

We found that principals in treatment schools did not, however, report having any more or less success recruiting teachers—in terms of interviews per vacancy or acceptances per offer made—than principals of control schools (Table V.8). Principals in control schools reported having about 1.5 more teacher vacancies per school than treatment principals, a difference that is consistent with a higher teacher attrition rate in control schools, assuming the treatment and control schools were of similar size. For the six evaluation districts that provided teacher counts at the point of random assignment, the treatment schools had an average of 35.9 teachers, and the control schools had an average of 36.1 teachers. The difference (0.2) was not statistically significant.

Principals' Staffing Decisions

Because pay-for-performance bonuses depend on students' achievement growth on standardized tests, principals in schools eligible for pay-for-performance 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

⁴² The magnitude of the impact on likelihood of using pay-for-performance as a recruiting incentive varies across model specifications (Appendix E, Table E.8). The impact of using the TIF program as a recruiting incentive is similar across model specifications.

principal may decide to assign teachers to tested grades and subjects based on belief in a teacher's ability to raise student achievement scores.

Table V.7. Incentives Used to Recruit Teachers (Percentages Who Reported They Were “Always” or “Often” Used)

Incentives Used for Recruiting Teachers	Treatment	Control	Impact
Salary	24.6	24.2	0.3
Opportunities to earn performance-based pay	26.4	16.7	9.8*
Opportunities for career advancement	25.8	23.1	2.7
Opportunities for professional development	61.5	62.1	-0.6
The level of teacher involvement in school decision making	53.0	59.1	-6.1
Collegiality of teaching staff	78.2	86.6	-8.3
The school culture and/or educational philosophy	84.0	86.6	-2.6
The school's reputation	73.2	71.6	1.5
The school's location or neighborhood	38.5	38.8	-0.3
The level of student achievement at the school	50.6	50.0	0.6
The TIF program	45.6	29.2	16.4*
Number of Principals—Range^a	64–67	65–67	

Source: Principal survey.

^aSample 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.8. Teaching Vacancies and Hiring Experiences (Averages Unless Otherwise Noted)

	Treatment	Control	Impact
Classrooms with teacher vacancies	3.0	4.5	-1.5*
Applications school reviewed for positions	27.3	27.9	-0.6
Applicants school interviewed	10.5	13.7	-3.1*
Offers school made	3.4	5.1	-1.7*
Offers that were accepted	3.0	4.6	-1.6*
Interview ratio (interviewed applicants divided by classroom vacancies)	3.8	4.1	-0.3
Acceptance ratio (offers accepted divided by offers made)	0.9	0.9	0.0
Number of Principals—Range^a	52–66	55–65	

Source: Principal survey.

^aSample 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.

For most measures of principals' staffing decisions, we found no significant impact of pay-for-performance. Although we found evidence that pay-for-performance resulted in principals using some different criteria to assign teachers to grades and subjects, the differences are not consistent with the hypothesis that pay-for-performance encouraged principals to strategically assign teachers to obtain pay-for-performance bonuses (Table V.9).⁴³ For example, when we focus on the criteria principals "always used" to assign teachers, principals in treatment schools were significantly less likely (by approximately 10 percentage points) to report using a teacher's ability to raise test scores (shown in Appendix E, Table E.5).

Table V.9. Criteria Used to Assign Teachers to Grade Levels or Subject Areas (Percentages Who Report They Are "Always" or "Often" Used)

	Treatment	Control	Impact
The teacher's experience in a grade level or subject area	89.5	88.1	1.4
The teacher's seniority	3.7	13.6	-9.9*
The teacher's content knowledge	92.6	96.9	-4.3
The teacher's ability to produce high test scores in grades/classes in which state or federal assessments are administered	75.3	75.4	-0.1
The teacher's ability to work with certain student populations	86.8	80.6	6.2
To balance teacher experience and expertise in a grade level or subject	73.0	70.8	2.3
Number of Principals—Range^a	67–68	65–67	

Source: Principal survey.

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

^aSample sizes are presented as a range based on the data available for each row in the table.

Pay-for-performance had some impacts on teachers' and principals' reported school preferences, but fewer than 6 percent of teachers reported that TIF affected their choice of school or subject area (Table V.10). Eligibility for pay-for-performance had some impact on teachers' school choice: more treatment than control school teachers reported that TIF affected their choice of school (5.5 versus 3.6 percent). Specifically, teachers in treatment schools were more likely to report that TIF influenced their decision to stay at or apply to their current school. This might suggest some recruitment and retention effects, although the effects are small.

Principals were more likely than teachers to report that the TIF program affected their school preference. Principals in treatment schools were more likely to report that TIF affected their choice of school (13 versus 9 percent) (see Table V.10). Specifically, principals in treatment schools were about 7 percentage points more likely to report that they stayed at their current school because of TIF (10 versus 3 percent).

Although less than 6 percent of teachers indicated that TIF affected their current choice of school, approximately one-fifth of teachers in treatment and control schools said that TIF will affect

⁴³ We found the same pattern of results when using a logit specification for this analysis (Appendix E, Table E.9).

their school preference for the coming year (Table V.11). About 13 percent of all teachers indicated that TIF would affect their desire to stay at their current school, but teachers in treatment schools were less likely to report that they planned to change schools to leave the TIF program (1.2 versus 3.2 percent). There were no other significant differences between treatment and control teachers' plans.⁴⁴

Table V.10. Influence of TIF Program on Educators' School Preference (Percentages)

	Treatment	Control	Impact
Teachers			
TIF Program Affected Where or What to Teach	5.5	3.6	1.9*
Ways in Which TIF Affected Where or What to Teach			
Stayed at school because of TIF	3.3	2.0	1.3*
Changed school to get into TIF	0.4	0.7	-0.3
Changed primary grade or subject because of TIF	0.5	0.8	-0.3
Applied to current school to get into TIF	1.8	0.8	1.1*
Applied for position in another school to leave TIF	0.0	0.2	-0.2
Applied for position in another school with better bonus program	0.0	0.0	0.0
Number of Teachers—Range^a	410–411	414	
Principals			
TIF program affected choice of school	12.9	9.1	3.8
Ways in Which TIF Affected School Preference			
Stayed at school because of TIF	9.7	3.0	6.7*
Came to school to get into TIF	3.1	6.1	-2.9
Number of Principals	68	66	

Source: Teacher and principal surveys.

^aSample 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.11. Influence of TIF Program on Teachers' School Preference for Next Year (Percentages)

	Treatment	Control	Impact
Teachers			
TIF Program Expected to Affect Preference of School for Next Year	21.0	18.8	2.2
Ways in Which TIF Program Will Affect School Preference			
Stay at current school because of TIF	13.7	11.7	2.0
Change school to get out of TIF	1.2	3.2	-2.1*
Change grade or subject because of TIF	1.4	1.2	0.2
Apply for position in another school to leave TIF	3.4	2.3	1.0
Apply for position in another school with better bonus program	2.2	1.4	0.8
Number of Teachers	411	414	

Source: Teacher surveys.

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

⁴⁴ Estimates are similar using alternate model specifications (see Appendix E, Table E.10).

We also examined the potential impact of pay-for-performance on teacher and principal mobility by examining the background characteristics (for example, race, age, teaching certification, and years of teaching experience) of teachers in treatment and control schools. Because schools were randomly assigned to the treatment or control group, any differences would suggest that certain types of teachers stayed in their schools or moved from their schools in order to take advantage of or avoid pay-for-performance. There were some statistically significant differences between treatment and control teachers and none between principals (Appendix E, Tables E.14–E.17). Teachers in schools that offered pay-for-performance were more likely to be white and less likely to be black. A higher percentage of teachers in treatment schools had majored in elementary education and had regular certification, and treatment school teachers were less likely to be in their first year of teaching in their current schools. Teachers in treatment schools were also less likely to have held a nonteaching job since college. Overall, no clear pattern emerged in the types of teacher who might have moved or remained in their current school in response to pay-for-performance.

The findings presented in this report are based on information collected from districts and educators during the 2011-2012 school year. Although the TIF program was implemented during that school year, feedback on some performance measures and bonuses based on performance measures were provided to educators after the school year ended. Thus the findings presented in this report are based on only part of the process playing out. Future reports will examine whether districts refined their TIF programs in response to their initial implementation experiences. In addition we may observe a change in educators' understanding of the TIF program, attitudes toward the program, or teaching strategies once they have experienced the complete feedback loop intended by the TIF grant generally, and pay-for-performance in particular. Future reports will also analyze the impact of pay-for-performance on educator mobility to see if educators respond to the pay-for-performance bonuses by changing the schools or subject areas in which they choose to teach. Finally, future reports will also examine whether there is an impact of pay-for-performance on student achievement.

REFERENCES

- Angrist, Joshua D., and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2009.
- Barron's Profiles of American Colleges 2003*. 25th ed. New York: Barron's Educational Series, 2002.
- Bayonas, Holli. "Guilford County Schools Mission Possible Program: Year 3 (2008–09) External Evaluation Report." Greensboro, NC: The SERVE Center, University of North Carolina at Greensboro, 2010.
- Dee, Thomas and James Wyckoff. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." Cambridge, MA: National Bureau of Economic Research, Working Paper No. 19529, October 2013.
- Fryer, Roland. *Teacher incentives and student achievement: Evidence from New York City public schools*. Cambridge, MA: National Bureau of Economic Research, Working Paper No. 16850, March 2011.
- Fryer, Roland, Steven Levitt, John List, and Sally Sadoff. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." National Bureau of Economic Research, Working Paper No. 18237, July 2012.
- 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, 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, and Allison Seifullah. "An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) After Four Years." Washington, DC: Mathematica Policy Research, March 2012.
- 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.
- Imai, Kosuke. "Variance Identification and Efficiency Analysis in Randomized Experiments Under the Matched-Pair Design." *Statistics in Medicine*, vol. 27, pp. 4857-4873. 2008.
- Imai, Kosuke, Gary King, and Clayton Nall. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science*, vol. 24, no. 1, pp. 29-53. 2009.

- Kamenica, Emir, “Behavioral Economics and Psychology of Incentives,” *Annual Review of Economics* Vol. 4, pp. 427-52. 2012.
- Liang, Kung-Yee, and Scott L. Zeger. “Longitudinal Data Analysis Using Generalized Linear Models.” *Biometrika*, vol. 73, pp. 13–22. 1986.
- Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., and Peng, A. *A big apple for educators: New York City’s experiment with schoolwide performance bonuses* (Final Evaluation Report). Santa Monica, CA: RAND Corporation. 2011.
- 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: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, October 2009.
- Rubin, Donald. *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley and Sons, 1987.
- Schafer, Joseph, and John Graham. “Missing Data: Our View of the State of the Art.” *Psychological Methods*, vol. 7, pp. 147-177, 2002.
- Schenker, Nathaniel, and Jeremy Taylor. “Partially Parametric Techniques for Multiple Imputation.” *Computation Statistics & Data Analysis*, vol. 22, pp. 425-446, 1996.
- Shifrer, Dara, Ruth Turley, and Holly Heard. “Houston Independent School District’s ASPIRE Program: Estimated Effects of Receiving Financial Awards.” Houston Education Research Consortium, Working Paper. 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.
- 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: National Center on Performance Incentives, Vanderbilt University, 2010.
- Springer, Matthew, Dale Ballou, and Art Peng. “Impact of the Teacher Advancement Program on Student Test Score Gains: Findings from an Independent Appraisal.” Working Paper 2008-19. Nashville, TN: National Center on Performance Incentives, Vanderbilt University, 2008.
- 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, 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. pp. 367-390. 2012.

THIS PAGE IS INTENTIONALLY BLANK

APPENDIX A

SUPPLEMENTARY INFORMATION ON STUDY SAMPLE AND DESIGN

THIS PAGE IS INTENTIONALLY BLANK

In this appendix, we provide more detailed information about the study sample and design. We discuss the procedures by which we randomly assigned schools to treatment and control groups within the evaluation districts, and we present descriptive statistics on the degree of similarity between these groups prior to the implementation of TIF. We also provide details on the methods used in sample selection for the teacher survey.

Characteristics of TIF Districts

In Chapter 2, we described the characteristics of TIF districts, focusing on the size, location, and students' socioeconomic status. More details about the characteristics of TIF districts compared with the average U.S. district are shown here, in Table A.1.

Table A.1. Comparison of TIF Districts to All U.S. Districts (Percentages Unless Otherwise Noted)

	All U.S. Districts	All TIF Districts
Student Racial/Ethnic Distribution		
White, non-Hispanic	69.0*	51.8
Black, non-Hispanic	11.8*	26.0
Hispanic	12.7*	18.4
Geographic Region		
Northeast	22.1*	9.2
Midwest	35.4*	28.1
South	22.0*	43.1
West	20.5*	19.6
Collective Bargaining ^a		
In state with collective bargaining	64.5*	35.9
Sample Sizes		
Number of districts	16,129	142
Number of states, including District of Columbia	51	25

Source: Common Core of Data for 2009–2010 school year.

Notes: Table is based on 142 of the 153 TIF districts that were included in the analyses, with 130 non-evaluation districts and 12 evaluation districts. Eleven non-evaluation districts were not included in the 2009–2010 district-level Common Core Data.

^aCollective bargaining is a state-level indicator from the National Right to Work Legal Defense Foundation (<http://www.nrtw.org/rtws.htm>).

^bSample sizes are presented as a range, based on the data available for each row in the table.

*Characteristics or distribution of characteristics are statistically different at the 0.05 level, based on a two-tailed test.

Random Assignment of Schools to 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 pairs of groups of schools.

Matched pairs of schools. We randomly assigned most of the schools (74 of 137) 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 pairs of groups of schools. For the remaining schools (63 of 137), 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.

Baseline School Characteristics

We tabulated the characteristics of study schools to understand the setting in which educators in the study worked, and to assess whether random assignment produced treatment and control groups that were equivalent at baseline (that is, prior to the implementation of TIF). We obtained data on school characteristics from the Common Core of Data and the U.S. Department of Education School-Level Assessment Data. In Table A.2, we show data pertaining to the 2009–2010 school year, which describes the schools before the random assignment that occurred between December 2010 and June 2011. As shown in the table, there were no statistically significant treatment–control differences in the school characteristics examined, including school size, percentage of students eligible for free or reduced-price lunch, and percentage of students proficient in math and reading, among other characteristics.

Selection of the Teacher Survey Sample

As discussed in Chapter II, we surveyed a subset of the teachers in all of the 137 study schools that were randomized before the 2011–2012 school year. 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 survey.

Teaching assignments targeted by the survey. The target population for the teacher survey consisted of 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 an imbalance in grades or subjects.

Table A.2. Average Baseline Characteristics of Treatment and Control Schools

	Treatment	Control	Difference
School Type (percentages)			
Charter School	14.9	14.9	0.0
Enrollment			
Total enrollment	522	531	-9
School Location (percentages)			
Urban	57.0	52.2	4.8
Suburban	24.1	28.4	-4.2
Town	5.5	7.5	-2.0
Rural	11.9	11.9	0.0
Difference in distribution is significant	—	—	No
Student Racial/Ethnic Distribution			
Percentage white, non-Hispanic	28.2	30.5	-2.3
Percentage black, non-Hispanic	44.3	42.6	1.7
Percentage Hispanic	22.9	22.2	0.8
Percentage Asian	2.0	2.5	-0.5
Percentage other race/ethnicity	1.2	1.0	0.2
School Socioeconomic Status (percentages)			
Eligible for Schoolwide Title 1	92.2	91.0	1.2
Student Achievement Proficiency Rate ^a			
Percentage proficient in mathematics	67.5	68.0	-0.5
Percentage proficient in reading/English language arts	66.1	65.9	0.2
Number of Schools	66	67	

Source: Common Core of Data (school-level) and U.S. Department of Education School-Level Assessment Data in Reading and Math for 2009–2010 school year.

Notes: Four TIF schools were not found in Common Core of Data or U.S. Department of Education School-Level Assessment Data for the 2009–2010 school year.

^aDefined as the percent of students tested who achieve at the state-defined level of proficiency or above. If a school reported student achievement proficiency as a range, the lower bound was used in calculating the reported estimate.

We chose the set of targeted 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 a number of reasons. First grade has full-day classes and is less likely to have standardized testing than grades 2 and 3. Science is a well-defined subject that is not tested annually, and retaining certified science teachers is an important policy goal.

Sampling methods. Within each study school, we used administrative data provided by the evaluation districts to identify teachers who were assigned to any of the grades and subjects in our target population. These teachers constituted our sampling frame.

Because our future analysis of impacts on student achievement will be focused on tested grades and subjects, our sampling approach for the teacher survey was also designed to give greater emphasis to tested grades and subjects relative to nontested ones. In each study school, we selected all teachers who taught any of the tested grades and subjects targeted by the survey. Additionally, we 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 set of 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.¹

Our initial sample included all teachers for whom the teacher assignment data provided by TIF districts indicated were or may be teaching the targeted grades and subjects. However, because some teaching rosters were not sufficiently detailed (for instance, describing teachers' grades as a range of grades) or were inaccurate, our sample included 97 teachers who reported in the survey that they were not teaching in the targeted grades and subjects. We excluded these teachers from the 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.

¹ Due to an error in the sampling algorithm, in three districts we inadvertently sampled all 1st grade teachers in the study schools.

APPENDIX B

SURVEY RESPONSE RATES AND CHARACTERISTICS OF RESPONDENTS

THIS PAGE IS INTENTIONALLY BLANK

Table B.1. District Survey Response Rates Overall and by Evaluation Status

	Overall	Evaluation Districts	Non-Evaluation Districts
All Districts			
Number of districts	182	14	168
Number of respondents	165	14	151
Number of ineligible respondents	12	2	10
Response rate (respondents over total)	90.7%	100%	89.9%

Source: District surveys.

Notes: Ineligible districts are districts that dropped out of TIF or were not implementing TIF at the time of survey administration.

*Response rate difference between non-evaluation and evaluation is statistically significant at the .05 level, two-tailed test.

Table B.2. District Characteristics by District's Response Status (Percentages Unless Otherwise Noted)

	Respondents	Respondents and Nonrespondents
Student Racial/Ethnic Distribution		
White, non-Hispanic	50.5*	48.5
Black, non-Hispanic	26.1	26.8
Hispanic	19.3	18.8
Asian	1.7	1.7
Other	1.5*	3.3
Student Socioeconomic Status		
Eligible for free/reduced-price lunch	63.3	64.2
Title 1 eligible schools (school-wide)	83.6	83.8
Enrollment and Staffing (averages)		
Total enrollment	28,730	31,394
Student/teacher ratio	15.9	15.9
District Location^a		
Urban	36.0	36.7
Suburban	14.7	15.1
Town	20.0	20.5
Rural	29.3	27.7
Difference in distribution is significant	No	--
District Census Bureau Region		
Northeast	6.7	6.6
Midwest	27.3	25.3
South	46.0	46.4
West	20.0	21.7
Difference in distribution is significant	No	--
Number of Districts	148	162

Source: Common Core of Data for 2009–2010 school year.

Notes: 16 TIF non-evaluation districts are not included in the 2009-2010 district-level Common Core Data. Title 1 eligible schools are calculated using school-level Common Core Data. All other demographic characteristics are calculated using district-level Common Core Data.

^aDistrict's location indicates the physical location of the district agency.

*Respondents significantly different from the sampled population at the 0.05 level, two-tailed test.

Table B.3. Teacher Respondents by Teaching Assignment and Treatment Status

Category	Treatment	Control	Total
Teachers			
1st grade only	121	137	258
4th grade only	137	133	270
7th grade English Language Arts and/or Math only	100	101	201
7th grade Science only	37	30	67
Other combination of eligible grades and subjects	16	14	30
Total	411	415	826

Source: Teacher survey.

Notes: Counts are for teachers eligible for the analysis. Appendix A describes the eligibility criteria for the sample.

Table B.4. Teacher and Principal Survey Response Rates Overall and by Treatment Status

Category	Overall	Treatment	Control	Difference
Teachers				
Number of teachers sampled	1,008	499	509	-10
Number of respondents	923	452	471	-19
Number of ineligible respondents	97	41	56	-15
Response rate (respondents over sampled)	91.6%	90.6%	92.5%	-1.9 %
Principals				
Number of principals sampled	138 ^a	69	69	0
Number of respondents	135	68	67	1
Number of ineligible respondents	0	0	0	0
Response rate (respondents over sampled)	97.8%	98.6%	97.1%	1.5 %

Source: Teacher and principal surveys.

Notes: Ineligible teachers and principals are respondents who work in a district not participating or implementing TIF in 2011-2012. In addition, part-time teachers or teachers for grades and subjects other than 1st, 4th, and 7th grade math, English Language Arts, and science are also considered ineligible.

Table B.5. Survey Response Rates, Teacher Survey (Percentages)

	Teacher Survey			
	Overall	Treatment	Control	Difference
All	91.6	90.6	92.5	-1.9
District				
A	94.3	94.6	94.1	0.5
B	83.8	81.6	86.1	-4.5
C	96.2	100.0	92.3	7.7
D	94.8	87.5	100.0	-12.5
E	88.1	86.7	89.5	-2.8
F	95.0	100.0	91.4	8.6
G	93.4	93.9	92.7	1.2
H	90.2	88.5	92.0	-3.5
I	95.8	93.3	98.9	-5.6*
J	91.2	92.5	90.0	2.5
Number of Teachers Sampled	1,008	499	509	

Source: Teacher survey.

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

Table B.6. Survey Response Rates, Principal Survey (Percentages)

	Principal Survey			
	Overall	Treatment	Control	Difference
All	97.8	98.6	97.1	1.5
District				
A	100.0	100.0	100.0	0.0
B	100.0	100.0	100.0	0.0
C	75.0	100.0	50.0	50.0
D	100.0	100.0	100.0	0.0
E	95.0	90.0	100.0	-10.0
F	100.0	100.0	100.0	0.0
G	100.0	100.0	100.0	0.0
H	100.0	100.0	100.0	0.0
I	95.5	100.0	90.9	9.1
J	100.0	100.0	100.0	0.0
Number of Principals Sampled	138	69	69	

Source: Principal survey.

Potential Nonresponse Bias

Because of the high response rate (91 percent in the district survey, 92 in the teacher survey, and 98 in the principal survey), the potential for nonresponse bias is minimal. Nonetheless, we assessed the extent to which the respondents are representative of the full population of respondents and nonrespondents. In tables B.7-B.9, we compare respondents to the full sample on dimensions such as district location and size (district survey) or on school characteristics such as demographic composition and achievement proficiency (educator surveys). Survey respondents are statistically indistinguishable from the full set of respondents and nonrespondents in the majority of the characteristics. Because response rates were high and respondents were similar to the full sample, we did not impute any missing values.

Table B.7. School and Student Characteristics of Teacher Survey Respondents and Full Teacher Sample (Percentages)

	Teacher Survey	
	Respondents	Respondents and Nonrespondents
School Type		
Regular school	79.1	80.1*
Charter school	20.6	19.7*
School Size		
Smaller (<=600 students)	63.8	63.9
Larger (>600 students)	36.2	36.1
School Location		
Urban	55.7	56.7*
Other (suburban, town, rural)	44.0	43.0*
Student Racial/Ethnicity Distribution		
Majority White, non-Hispanic	26.1	25.4
Majority Black, non-Hispanic	48.3	48.8
Majority Hispanic	17.3	17.3
No majority	8.9	9.1
Student Poverty		
Lower poverty (<=80% Free or Reduced Price Lunch)	43.8	43.2
Higher poverty (> 80% Free or Reduced Price Lunch)	56.3	56.8
Student Achievement Proficiency Rate ^a		
High rate, Mathematics (>=80%)	23.3	23.3
Low rate, Mathematics (<80%)	76.7	76.7
High rate, Reading/English Language Arts (>=80%)	23.3	22.9
Low rate, Reading/English Language Arts (<80%)	76.7	77.1
Number of Teachers	893	977

Source: Common Core of Data (School-level) and U.S. Department of Education School-level Assessment Data in Reading and Math for 2009–2010 school year.

Notes: 4 TIF schools were not found in Common Core of Data or U.S. Department of Education School-level Assessment Data for 2009–2010 school year.

^aDefined as the percent of students tested who achieve at the state-defined grade-level of proficiency or above. If school reported student achievement proficiency as a range, lower bound is used in calculating the reported estimate.

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

Table B.8. School and Student Characteristics of Principal Survey Respondents and Full Principal Sample (Percentages)

	Principal Survey	
	Respondents	Respondents and Nonrespondents
School Type		
Regular School	83.3	83.0
Charter School	16.0	16.4
School Size		
Smaller (<=600 students)	70.5	69.6
Larger (>600 students)	29.5	30.4
School Location		
Urban	54.5	54.1
Other (suburban, town, rural)	44.7	45.2
Student Racial/Ethnicity Distribution		
Majority White, non-Hispanic	28.8	28.1
Majority Black, non-Hispanic	47.7	48.1
Majority Hispanic	15.9	16.3
No majority	9.1	8.9
Student Poverty		
Lower poverty (<=80% Free or Reduced Price Lunch)	43.9	43.7
Higher poverty (> 80% Free or Reduced Price Lunch)	56.1	56.3
Student Achievement Proficiency Rate ^a		
High rate, Mathematics (>=80%)	27.3	26.7
Low rate, Mathematics (<80%)	72.7	73.3
High rate, Reading/English Language Arts (>=80%)	27.3	26.7
Low rate, Reading/English Language Arts (<80%)	72.7	73.3
Number of Principals	131	134

Source: Common Core of Data (School-level) and U.S. Department of Education School-level Assessment Data in Reading and Math for 2009–2010 school year.

Notes: 4 TIF schools were not found in Common Core of Data or U.S. Department of Education School-level Assessment Data for 2009–2010 school year.

^aDefined as the percent of students tested who achieve at the state-defined grade-level of proficiency or above. If school reported student achievement proficiency as a range, lower bound is used in calculating the reported estimate.

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

Table B.9. District Reported Characteristics by District's Response Status on the Survey Question about the Distribution of Pay-for-Performance Bonuses for Teachers (Percentages of Districts)

	Pay-for-Performance Bonus Respondents	Pay-for-Performance Bonus Nonrespondents
Student Achievement Measures Used to Evaluate Teachers		
Any achievement measure	85.1	81.0
Achievement growth	85.1	81.0
By school	78.2	73.0
By student subgroups ^a	53.5	41.3
By teacher's classroom	74.7	61.9
Achievement level	49.4	39.7
Teacher Advancement Program (TAP) program	44.8*	27.3
Educators Could Earn Pay-for-Performance Bonuses Before	36.9	27.7
Percent of Districts that Revised TIF Program After Grant Award	54.1	64.1
To address budget limitations	27.1	35.9
To obtain the support of educators	25.3	27.0
Based on results of analysis of educator performance metrics	12.9	23.4
To better align with data-management systems	9.5*	21.9
Number of Districts	83–87	63–66

Source: District survey.

Notes: Districts are categorized based on whether they responded to the following survey question: "Among teachers who are eligible for performance-based bonuses or awards, what percentage of teachers in tested grades and subjects are expected to receive additional pay based on their performance this school year? a. \$0; b. \$1-\$999; c. \$1,000-\$1,999; d. \$2,000-\$3,999; e. \$4,000-\$5,999; f. \$6,000-\$7,999; g. \$8,000-\$9,999; h. \$10,000-\$11,999; i. \$12,000-\$14,999; j. \$15,000 or more."

^aStudent subgroups can be defined by grade, teams, subject areas, and demographic characteristics.

*Respondents significantly different from nonrespondents at the 0.05 level, two-tailed test.

THIS PAGE IS INTENTIONALLY BLANK

APPENDIX C

ANALYTIC METHODS AND SENSITIVITY ANALYSES

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 discuss the rationale for using districts, rather than grantees, as the unit of analysis when describing responses to district surveys regarding TIF implementation. Second, we describe how we constructed our measure of the maximum pay-for-performance bonus amount reported by each district. Third, we provide details of the analytic methods used to estimate impacts of pay-for-performance on the self-reported outcomes of teachers and principals. 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.

Rationale for Districts as the Unit for Analyzing District Survey Responses

Even though many grantees included multiple districts, the analyses use districts rather than grantees as the unit of analysis for two reasons. First, in our role as the technical assistance provider for evaluation grantees, we learned that within multidistrict grantees there were important differences in the programs districts implemented. Second, we conducted an analysis that verified that there was considerable variation in programs within grantees. In particular, for key outcomes, we calculated the grantee-level intraclass correlation coefficient (ICC)—the proportion of the total variance of the outcome that was observed to occur across grantees—among districts that belonged to multidistrict grantees. As shown in Table C.1, for most key outcomes, most of the variance occurred within grantees rather than between grantees, as evidenced by ICC values lower than 0.5.

Table C.1. Interclass Correlation in Multidistrict Grantees

	Interclass Correlation Coefficient	Number of Districts in Multidistrict Grantees	Number of Multidistrict Grantees
Maximum Pay-for-Performance Bonus Possible ^a	0.32	107	12
Pay-for-Performance Bonus for Teachers in Tested Grades and Subjects			
Average	0.39	60	11
Maximum	0.45	60	11
Minimum	0.36	60	11
Pay-for-Performance Bonus for Teachers in Nontested Grades and Subjects			
Average	0.43	60	11
Maximum	0.29	60	11
Minimum	0.19	60	11
Pay-for-Performance Bonus for Principals			
Average	0.54	72	11
Maximum	0.63	72	11
Minimum	0.49	72	11
Average Number of Classroom Observations per School Year	0.00	105	12

Source: District survey and district interviews.

Note: Table is based only on grantees working with more than one district.

^aPay-for-performance bonus is calculated using a combination of survey questions and district interviews as described in this appendix.

Construct for Measuring Districts' Reports of Maximum Performance Bonus Amounts

In Chapters III and IV, we discussed districts' reports of the average maximum pay-for-performance bonus amounts for which teachers and principals were eligible. Here, we describe the way in which we constructed our main measure of the district-reported maximum bonus amounts. We also describe an alternative measure that we used in some analyses because of missing data in the main measure.

Main measure. The main measure of district-reported maximum pay-for-performance bonus amounts was based on survey questions in which districts were asked to report the expected distribution of pay-for-performance bonuses for educators in their TIF schools. Districts reported expected bonus distributions separately for each of three categories of educators—teachers in tested grades and subjects, teachers in nontested grades and subjects, and principals. Specifically, districts were asked to report the percentage of educators who were expected to receive bonuses in each of the following intervals: (1) \$0; (2) \$1–\$999; (3) \$1,000–\$1,999; (4) \$2,000–\$3,999; (5) \$4,000–\$5,999; (6) \$6,000–\$7,999; (7) \$8,000–\$9,999; (8) \$10,000–\$11,999; (9) \$12,000–\$14,999; and (10) \$15,000 or more. For each district, we calculated the maximum pay-for-performance bonus amount as the upper limit of the highest interval in which a positive percentage of educators was expected to receive a bonus.¹ We used this approach in the main analysis because of its simplicity and transparency, given that it was based on a single survey question pertaining to each type of educator.

Alternative measure. In one specific type of analysis—estimates of the impacts of pay-for-performance within subgroups of evaluation districts defined by the size of teachers' maximum pay-for-performance bonus amount—we needed information on the maximum pay-for-performance bonus amount in every evaluation district. However, only 5 of the 10 evaluation districts in the impact analysis answered the relevant question needed to construct the main measure of district-reported bonus amounts. Therefore, for this analysis, we used an alternative measure that drew upon a combination of district survey items and interview questions—an approach that yielded no missing values for evaluation districts.

The alternative approach used two additional questions from the district survey, labeled below by their item numbers (D6 and D7) from the survey:

- D6: What is the maximum amount of additional pay that a teacher in any of your TIF schools could receive because of his or her overall performance this school year?
- D7: For which of the following performance criteria are teachers in tested grades and subjects eligible to receive additional pay this school year in any of your TIF schools? For each yes answer, write the maximum amount that a teacher could receive.
 - (a) Student achievement level
 - (b) Student achievement growth at the school level
 - (c) Student achievement growth in certain student groups
 - (d) Student achievement growth in teacher's individual classes
 - (e) Other measures.

¹ The upper limit of the highest possible interval was assumed to be \$15,000.

Using a district’s responses to these questions, we constructed a measure of the maximum pay-for-performance bonus amount available to teachers based on the following steps. First, the bonus amount equaled the district’s response to question D6 if it was nonmissing. Second, if the response to question D6 was missing, the sum of the maximum bonus amounts for the different performance criteria listed in question D7 was used, but only if that sum was consistent with the expected distribution of bonuses for tested grades and subjects from the main approach described earlier—that is, if the sum fell in the highest interval of bonus amounts that was expected to have a positive percentage of teachers. If the sum was not consistent with the main approach, the amount based on the main approach was used.

For evaluation districts, we also asked about maximum pay-for-performance bonuses in our in-depth phone interviews with TIF program coordinators. If the interview response was substantially different from the one provided in the survey (or if relevant survey questions were left missing), the interview response was used instead.

Estimating Impacts of Pay-for-Performance on Teacher- and Principal-Reported Outcomes

Below, we describe the estimation model we used to estimate impacts of pay-for-performance on intermediate outcomes reported by teachers and principals, which we presented in Chapter V. We then discuss details of the sensitivity analyses and subgroup analyses that we conducted as part of the experimental analysis. For expositional simplicity, we refer primarily to impacts on intermediate outcomes, but the same analytic methods were used to estimate treatment-control differences in educators’ reports of TIF implementation, which we presented in Chapter IV.

Main Estimation Model

In the experimental analyses, the key parameter of interest was the impact of schools’ eligibility for pay-for-performance on intermediate outcomes. To estimate these impacts, we used a regression model that reflects the random-assignment design—specifically, the assignment of clusters of educators rather than individual educators, and the pairing of these clusters before random assignment.

To assess treatment-control differences in intermediate outcomes reported by teachers or principals, we estimated the following model:

$$(1) \quad y_{ispd} = \alpha_{pd} + \beta_d T_{spd} + \varepsilon_{ispd},$$

where y_{ispd} is the intermediate outcome reported by teacher or principal i in school s within matched pair p in district d , α_{pd} is a fixed effect for the randomization block (matched pair of schools or pair of groupings of schools), T_{spd} is an indicator equal to 1 if school s was assigned to the treatment group and equal to zero otherwise, and ε_{ispd} is a random error term.

We weighted teacher responses so that each school contributed equally to the average impact estimate. Specifically, we assigned weights to teachers with nonmissing outcomes so that the sum of their weights was equal across all schools. (No weights were used for principal analyses because there was only one principal per school.) A teacher j in school s was weighted by weight $W_{js} = 1/N_s$, where N_s is the number of teachers with nonmissing values for the outcome in school s . By applying

the same weighting approach to analyses of student-level achievement outcomes in the second report, we will be able to compare the findings from the second report to those from this report.

To calculate the average impact of pay-for-performance for the full study sample, we took a weighted average of the district-specific impact estimates from equation (1), $\hat{\beta}_d$, with each district weighted by the number of schools in the evaluation. We gave greater weight to districts with more participating schools because they contributed more information about impacts.

We estimated equation (1) using ordinary least squares (OLS) and employed Huber-White sandwich standard errors (Liang and Zeger 1986) to account for the clustering of teacher and principal outcomes at the level of the random-assignment unit (schools or groups of schools). These standard errors are robust to any arbitrary form of correlation among outcomes in the same cluster.

Because the main goal of the analysis was to gauge the overall impact of the program, outcomes were defined for the entire sample with nonmissing responses. That is, educators who reported not being eligible for pay-for-performance bonuses were considered to report a zero dollar bonus amount when pay-for-performance bonus amounts were the outcome.

Because the outcomes in this analysis were not the final outcomes of greatest importance to policy and of greatest interest to the evaluation (which are student achievement and educator mobility), we did not adjust hypothesis tests for multiple hypotheses. Given that we estimated impacts on several intermediate outcomes, the probability of erroneously finding a statistically significant impact at the 5 percent level, in the absence of a true impact, was greater than 5 percent.

Sensitivity Analyses

We assessed the sensitivity of the impact findings to a variety of model specifications.

Weights. The main report presented findings that assigned equal weights to schools and that weighted district-specific impacts based on their share of TIF schools. To assess robustness, we also estimated impacts in a variety of alternative ways. First, we used no teacher-level weights (implicitly weighting schools by the number of teacher respondents). Second, we weighted districts equally when combining the district-specific impacts. Third, we estimated a simplified model that directly estimated one overall impact across all districts.

Standard errors. In addition to the primary analysis approach for calculating standard errors presented in the body of the report, we used two alternative methods. First, we made less-conservative assumptions about the clustering of outcomes within certain types of assignment groups. In the main approach, we allowed for the possibility that outcomes from *different* schools in the *same* assignment group may be correlated. We might expect such correlations in cases where schools with a common feature—for instance, the same feeder pattern or the same charter management organization—were explicitly grouped together into the same assignment group. However, in two particular districts, the entire set of participating schools was divided into two assignment groups that were similar to each other on baseline characteristics. In these two districts, there was no compelling reason to suspect that different schools in the same assignment group should have correlated outcomes. Therefore, our sensitivity analysis treated schools—not assignment groups—in these two districts as the level of clustering used to calculate robust standard errors.

In our second sensitivity analysis, we calculated standard errors based solely on the randomization design. The robust standard errors in the main analyses were derived from the particular statistical model used to estimate impacts. Some researchers (see Imai 2008; Imai et al. 2009) advocate imposing no model on experimental data. They instead calculate variances of impact estimators based only on the probability distribution of the treatment assignment variable—in other words, based solely on how much the impact estimates would fluctuate if sample members were repeatedly reassigned between the treatment and control groups. We employed Imai et al. (2009) formulas for randomization-based impact estimates and standard errors based on random assignment conducted within matched pairs of clusters. (Randomization-based point estimates were slightly different than the main estimates because the randomization approach assigned an explicit weight to each block equal to the number of schools, whereas the main approach did not. Nevertheless, the key differences of interest in this sensitivity analysis were those reflected in the standard errors.)

Binary outcomes. Following Angrist and Pischke (2009), we reported OLS estimates from equation (1) for the experimental impact estimates regardless of whether the outcome variable was continuous or discrete. The OLS estimates have a clear, straightforward interpretation that is consistent with the experimental design: starting from the within-pair differences in mean outcomes between treatment and control schools, each district-specific impact estimate is simply a weighted average of these differences across all pairs in the district, with pairs weighted by the number of schools. Nevertheless, to assess the sensitivity of findings on binary outcomes, we also estimated a variant of equation (1) using a logit model. We report the marginal effects in Appendix E.

In theory, logit models assumed a more realistic functional form for the relationship between binary outcomes and the independent variables. However, a notable shortcoming of the logit models was that they could not reflect precision gains from the blocked random-assignment design. Accounting for these precision gains would have required the inclusion of randomization block indicators in the logit model. However, this would have led the model to drop all blocks in which the outcome variable did not vary, resulting in a final analysis sample that would not necessarily be representative of the original analysis sample. Therefore, we excluded randomization block indicators from the set of covariates; instead, to ensure that treatment status was not confounded with blocks, we reweighted the data so that the total weight assigned to treatment and control group members within the same block was equal. Because the logit model did not maximize precision, we used this model only to compare the magnitudes and direction of the impact estimates with those from our linear (primary) model, but we continued to use significance tests exclusively from the linear model—the most precise specification.

Estimation Model for Subgroup Analyses

We conducted analyses separately by subgroups to assess how the impacts of pay-for-performance differed by teachers' teaching assignment, teaching experience, or districts' program characteristics. 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 $Group1_{ispdb}$, $Group2_{ispdb}$, and $Group3_{ispdb}$ respectively. We estimated the following estimation model:

$$(2) \quad y_{ispd} = \alpha_{pd} + \beta_1 T_{spd} + \delta_1 Group2_{ispd} + \delta_2 Group3_{ispd} \\ + \beta_2 (T_{spd} \times Group2_{ispd}) + \beta_3 (T_{spd} \times Group3_{ispd}) + \varepsilon_{ispd}.$$

In equation (2), β_2 and β_3 capture the impact of pay-for-performance on teachers in groups 2 and 3 relative to the impact in group 1. We tested the statistical significance of the estimates of β_2 and β_3 to determine whether impacts differed across subgroups. For scenarios in which teachers were partitioned into two (rather than three) subgroups, equation (2) was identical except that it did not include indicators and interaction terms involving $Group3_{ispd}$.

Method for Imputing Missing Values of Educator-Reported Bonus Amounts

For nearly all analyses examining educators' reports of TIF implementation and intermediate outcomes, we assumed that respondents' reports were representative of the reports that the full sample of respondents and nonrespondents would have provided if everyone had answered the relevant survey items. Specifically, for teacher-survey items, the respondents from each school received, in total, the full weight (of one) assigned to that school, as described earlier. This approach assumed that respondents and nonrespondents from the same school did not differ systematically in the answers they would have given on an item. For treatment-control comparisons, this approach assumed only that any respondent-nonrespondent differences did not differ systematically across treatment groups.

For principal-survey items, our analysis used the unweighted sample of respondents to estimate equation (1) under the assumption that, after controlling for randomization block and treatment status, respondents' experiences did not differ systematically from those of nonrespondents. For both teachers and principals, our analyses in Appendix B, which found few differences in observed characteristics between respondents and the full sample, provide some support for our assumption that respondent' experiences could represent those of the full sample.

For one set of survey items—those that asked educators to report the maximum bonus amounts for which they or their same-school colleagues were eligible—we used a different approach to handling missing data because the occurrence of nonresponse depended upon a specific outcome: 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 or their colleagues 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 composed of teachers who reported being ineligible, and had a maximum bonus amount of zero. As a result, 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 the following five steps. First, we estimated an imputation model in which the reported maximum bonus amount was modeled as a linear function of treatment status and randomization block indicators—the same model as Equation (1). We estimated the imputation model using only teachers who reported being eligible for the specified bonus *and* reported a nonmissing bonus amount.² Second, the estimated coefficients and standard errors from the imputation model were used to form a posterior distribution 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 the three respondents, and the reported maximum bonus amount of the selected respondent served as the imputed value for the nonrespondent.³

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 (1), 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 (1).

We used the same approach to impute principal-reported maximum bonus amounts. However, unlike in the case of teachers, we did not control for randomization block indicators in the imputation model due to the small number of principal respondents per block. Instead, we controlled for district indicators.

² 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 in order 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.

³ 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.

THIS PAGE IS INTENTIONALLY BLANK

APPENDIX D

**SUPPLEMENTAL FINDINGS ON TIF DESIGN AND IMPLEMENTATION
FOR CHAPTERS III AND IV**

THIS PAGE IS INTENTIONALLY BLANK

To supplement the findings presented in Chapters III and IV, we present in this appendix additional analyses of TIF districts and their programs, and implementation of TIF in the evaluation districts.

TIF Districts and Their Programs

In this section, we provide additional detail for the findings presented in Chapter III on the design of TIF in all TIF districts. We described in that chapter the percentage of TIF districts that implemented the four required components of TIF. The TIF application notice also required that grantees implement five core elements to support implementation of their performance-based compensation system. One of those core elements—a rigorous, transparent, and fair evaluation system—required evidence in addition to student achievement and classroom observations to evaluate educators. Districts may have incorporated this evidence as part of their observation measure (for example, a principal’s professional judgment), or used the evidence as a separate measure (for example, parent or student input). In Table D.1, we show the types of additional evidence that TIF districts used to evaluate teachers. Among them: principal’s or other administrator’s judgment (69 percent of TIF districts), and teacher participation in school activities (41 percent).

Table D.1. Percent of Districts Using Additional Evidence To Evaluate Teachers and Principals

	All TIF Districts
Additional Evidence to Evaluate Teachers	
Principal’s or other administrator’s professional judgment	68.7
Teacher participation in school activities	40.8
Teacher attendance	34.0
Reviews from other teachers	13.8
Parent or student input	11.9
Student attendance	9.2
Additional Evidence to Evaluate Principals	
Teacher assessments of principal performance	47.7
Parent input	14.6
Number of Districts—Range^a	150–153

Source: District survey.

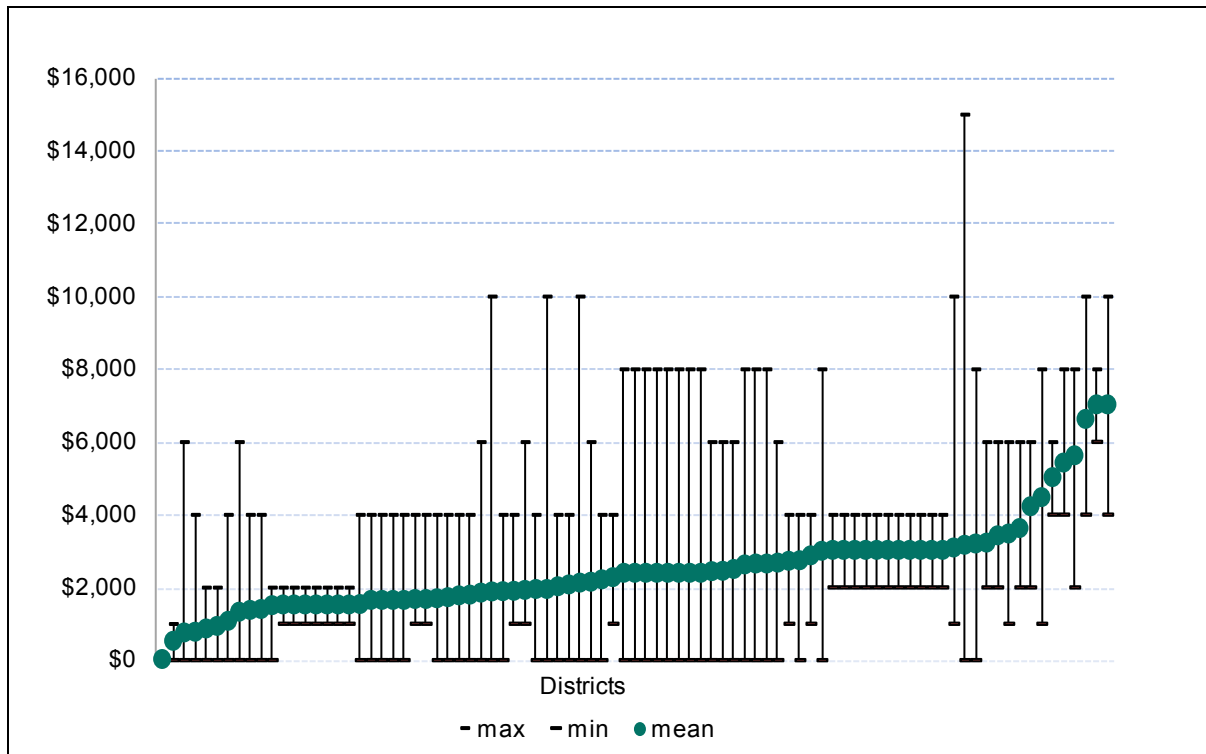
Note: Sample size may vary for individual items due to item nonresponse. The minimum and maximum sample size are displayed in the table.

^aSample sizes are presented as a range, based on the data available for each row in the table.

We presented information on the expected size of pay-for-performance bonuses in Chapter III, including the minimum, maximum, and average pay-for-performance bonus that TIF districts expected to pay out, and that information—provided as averages across all TIF districts—was shown in Figure III.1. In Figure D.1, we present the same information for each district separately, showing the distribution of expected minimum, maximum, and average pay-for-performance bonuses across districts. Each line in the figure represents a district and the top of the line is the maximum expected pay-for-performance bonus, the bottom of the line the minimum expected pay-for-performance bonus, and the green circle the average expected pay-for-performance bonus, with all amounts pertaining to teachers in tested grades and subjects. As noted in Chapter III, there is variation across districts in the range of expected pay-for-performance bonuses. For example, some districts have a difference of \$1,000 between the minimum and maximum expected pay-for-performance bonus; others have a difference of \$8,000 or more. In addition, the minimum expected

pay-for-performance bonus in some districts is at least \$2,000 in some districts, and as high as \$6,000 in one district.

Figure D.1. Pay-for-Performance Bonuses for Teachers in Tested Grades and Subjects, by Districts



Source: District survey.

Note: 6 evaluation and 81 non-evaluation districts responded to this survey question. The maximum, minimum, and average expected teacher performance bonus by district is shown in the figure. Districts are sorted by their mean value. Figure is based on a survey question about the expected distribution of TIF pay-for-performance bonuses, given 10 categories of bonus amounts that range from \$0 to \$15,000 or more (for example, the percentage of principals expected to earn a bonus between \$1,000 and \$1,999). The maximum and minimum values were calculated as the upper and lower range of the highest or lowest category with a positive percentage of teachers. The average is calculated as the sum of the midpoint of the amount category weighted by percentage of teachers expected to receive a bonus in a given category.

In Chapter III, we presented information on the percentage of TIF districts offering additional pay opportunities for teachers. We show districts’ past experience providing additional pay opportunities in Table D.2. Although one-third of TIF districts had past experience providing performance pay, 53 percent had past experience with additional pay opportunities. In Table D.3, we describe the magnitude of additional pay opportunities currently offered by TIF districts as part of their grant, and we compare each pay opportunity to the maximum pay-for-performance bonus offered on average across all TIF districts. The first two columns show the percentage of districts offering an opportunity for additional pay and the maximum amount that teachers could earn for each pay opportunity, on average across those districts. The next column shows that maximum amount as a percent of the maximum pay-for-performance bonus offered across all TIF districts. As a result, the sample of districts used to calculate the average maximum pay-for-performance bonus can differ from those used to calculate the maximum amount of a given pay opportunity.

Table D.2. Districts’ Past Experience Providing Pay-for-Performance Bonuses or Opportunities for Additional Pay (Percentages)

	Teachers	Principals
Educators Could Earn Pay-for-Performance Bonuses	32.9	28.6
Educators Could Earn Additional Pay for		
Professional development or graduate-level courses	36.2	17.1
Additional responsibilities	53.0	16.5
Teaching high-need subjects	18.1	-
Working in hard-to-staff school	13.4	11.8
Number of Districts—Range^a	149	139–144

Source: District survey.

^aSample sizes are presented as a range, based on the data available for each row in the table.

Table D.3. Additional Pay Opportunities for Teachers, Compared with Pay-for-Performance Bonus

Opportunity for Additional Pay for Teachers	Percent of TIF Districts That Offer Additional Pay (Percentage)	Districts That Offer the Specified Opportunity		
		Maximum Pay for Opportunity (Dollars)	Maximum Pay for Opportunity as a Percentage of the Average Maximum Pay-for-Performance Bonus ^a (Percentage)	Number of Districts That Reported Pay for Opportunity
Teachers Could Receive Additional Pay for Taking on Added Roles or Responsibilities	86.6	n.a.	n.a.	
Roles and responsibilities				
Mentor teacher	66.2	\$3,735	69.7	88
Master or lead teacher	55.1	\$7,145	133.4	72
Department chair or head	22.3	\$1,416	26.4	26
Lead curriculum specialist	8.9	\$2,320	43.3	10
Serving on a schoolwide committee or task force	16.9	\$1,256	23.5	20
Leadership team member	23.4	\$1,107	20.7	28
Additional Factors				
Teaching in hard-to-staff school	17.4	\$3,602	67.3	21
Teaching in high-need subjects	23.6	\$3,455	64.5	32
Attending professional development activities or enrolling in graduate-level courses	27.8	\$780	14.6	31
Number of Districts—Range^b	144–149	10–88		

Source: District survey.

Note: Table reports on activities funded by TIF.

^aThe average maximum pay-for-performance bonus for teachers in tested grades and subjects across all TIF districts is \$5,355 (see Figure III.1). This amount serves only as a descriptive benchmark because it is based on all districts that answered the pay-for-performance amount question and is not the same sample of districts that reported additional pay opportunities.

^bSample sizes are presented as a range based on the data available for each row in the table.

n.a.= not applicable.

The additional pay opportunities ranged from 15 percent to 133 percent of the average maximum pay-for-performance bonus expected across districts. Additional pay for master or lead teachers was the only opportunity that exceeded the average maximum pay-for-performance bonus expected across districts. The most common pay opportunity—additional pay for serving as a mentor teacher—represented 70 percent of the average maximum pay-for-performance bonus. The only additional pay opportunity that exceeded the average maximum pay-for-performance bonus amount was the master or lead teacher role, which was offered in 55 percent of districts.

TIF Implementation in Evaluation Districts

In this section, we provide more detailed findings about the design and implementation of TIF programs in evaluation districts to supplement the information presented in Chapter IV.

Design of TIF Programs in Evaluation Districts

In Chapter IV, we described the TIF programs implemented by evaluation districts with a focus on the four required TIF program elements. Here, we provide more details about the TIF program design in evaluation districts.

We showed how evaluation districts were more likely to evaluate teachers in TIF schools based on two classroom observations. In Table D.4, we show that evaluation and non-evaluation districts did not differ in their use of classroom observations to evaluate teachers. Differences in the percent of districts conducting classroom observations, the number of observations per school year, the length of observations, and the types of staff conducting observations were not significant across evaluation and non-evaluation districts.

As mentioned above, as part of the required core elements to support implementation, TIF districts had to use of additional forms of evidence—beyond student achievement and classroom observations—to evaluate educators. Overall, 81 percent of TIF districts used at least one additional type of evidence to evaluate teachers. Evaluation and non-evaluation districts did not differ in their use of additional evidence (Table D.4). In both types of districts, the most common type of additional evidence was a principal's or other administrator's professional judgment, followed by teacher participation in school activities and teacher attendance. The other types of evidence were used by less than 20 percent of evaluation and non-evaluation districts. Sixty percent of districts used at least one additional type of evidence to evaluate principals, with no statistically significant differences between evaluation and non-evaluation districts. The most common type of additional evidence used to evaluate principals was teacher assessments of principal performance, used by 50 percent of evaluation districts and 48 percent of non-evaluation districts. There were no statistically significant differences in the percent of evaluation and non-evaluation districts using other types of evidence to evaluate principals (Table D.5).

Table D.4. Districts’ Reports of Teacher Evaluation Measures (Percentage Unless Otherwise Indicated)

Teacher Performance Measure	Evaluation Districts	Non-Evaluation Districts
Classroom Observations		
Conduct classroom observations	100.0	97.8
Average number of observations per school year	3.3	4.7
Average length of observations in minutes	39.8	43.1
Districts in which observations are conducted by:		
Principals or other administrators at the teacher’s school	83.3	96.4
Teacher leaders or peer observers	66.7	53.2
Content specialists	33.3	17.4
District-level staff	0.0	8.7
Externally hired observers	16.7	2.9
Districts in which observations are conducted by principal or other administrators at the teacher’s school only	16.7	29.0
Additional Evidence for Teacher Evaluations		
Principal’s or other administrator’s professional judgment	66.7	68.8
Teacher participation in school activities	33.3	41.4
Teacher attendance	25.0	34.8
Reviews from other teachers	16.7	13.6
Parent or student input	8.3	12.2
Student attendance	8.3	9.3
At least one additional measure	75.0	81.2
Number of Districts—Range^a	12	134–141

Source: District survey.

Note: Sample size may vary for individual items due to item nonresponse. The minimum and maximum sample size are displayed in the table.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

Table D.5. District Report About Additional Evidence Used for Principal Evaluations (Percentages)

Additional Evidence for Evaluating Principals	Evaluation Districts	Non-Evaluation Districts
Teacher assessments of principal performance	50.0	47.5
Parent input	9.1	15.0
At least one type of additional evidence	58.3	60.3
Number of Districts—Range^a	11–12	140–141

Source: District survey.

Note: Sample size may vary for individual items due to item nonresponse. The minimum and maximum sample size are displayed in the table.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

In Chapter IV, we showed that evaluation districts were more likely to offer additional pay for roles and responsibilities, and we examined the magnitude of these additional pay opportunities. More detailed information on the magnitude of these pay opportunities is in Table D.6. We compare the size of these pay opportunities (for districts that offered the opportunity) to the size of the maximum pay-for-performance bonuses (for all districts). The additional pay opportunities offered by non-evaluation districts represent 16 to 144 percent of the size of pay-for-performance bonuses; those offered by evaluation districts represent 7 to 60 percent of the size of pay-for-performance bonuses. Fifty-four percent of non-evaluation districts and 73 percent of evaluation districts offered additional pay for serving as a master or lead teachers, which was the most common type of additional pay opportunity offered by TIF districts.

Teacher and Principal Perspectives on TIF Implementation

In this section, we present additional detail about the findings in Chapter IV related to teacher and principal perspectives on TIF implementation. First, we compare educators' and districts' reports of additional TIF program features that were not required under the grant rules. Second, we present findings on teachers' participation in professional development. Third, we discuss implementation of selected TIF components as reported by key subgroups of teachers. Finally, we describe educators' reports of maximum pay-for-performance bonus amounts without the imputed values that were included in the main analyses in Chapter IV.

Additional TIF program features. In Chapter IV, we compared educators' and districts' reports of whether two required types of performance measures—student achievement growth and classroom observations—were being used to evaluate teachers. In Table D.7 we describe educators' and districts' reports of additional evidence used to evaluate teachers, as required under the core elements needed to support TIF implementation. Teachers were less likely than their principals to report that teachers were evaluated based on participation in school activities, but were more likely to report that reviews from other teachers and student attendance factored into teacher evaluations. Districts' reports did not differ by a statistically significant margin from those of teachers or principals.

Table D.6. Comparison of Additional Pay Opportunities and Pay-for-Performance Bonuses, by Evaluation Status

	Evaluation Districts			Non-Evaluation Districts		
	Percent That Offered Additional Pay (Percentage)	Maximum Pay for Specified Opportunity (Dollars)	Maximum Pay for Opportunity as a Percent of the Average Maximum Pay-for-Performance Bonus ^b (Percentage)	Percent That Offered Additional Pay (Percentage)	Maximum Pay for Specified Opportunity (Dollars)	Maximum Pay for Opportunity as a Percent of the Average Maximum Pay-for-Performance Bonus ^b (Percentage)
Opportunity for Additional Pay for Teachers						
Teachers Could Receive Additional Pay for Taking on Added Roles or Responsibilities	100.0*	n.a.	n.a.	85.5	n.a.	n.a.
Roles and responsibilities						
Mentor teacher	90.9*	\$3,460	40.7	64.2	\$3,770	73.6
Master or lead teacher	72.7	\$5,104	60.1	53.7	\$7,400	144.5
Support school-, grade-, or subject-level decisions ^a	50.0	\$2,542	29.9	39.0	\$1,495	29.2
Additional Factors						
Teaching in hard-to-staff school or high-need subject	33.3	\$4,725	55.6	30.3	\$3,518	68.7
Attending professional development activities or enrolling in graduate-level courses	33.3	\$633	7.4	27.3	\$796	15.5
Number of Districts—Range^c	11–12	3–10	3–10	132–141	28–78	28–78

Source: District survey.

Note: Table reports on activities funded by TIF.

^aThis includes being a department chair, a lead curriculum specialist, or serving on a schoolwide committee or as a leadership team member.

^bThe average maximum pay-for-performance bonus for teachers in tested grades and subjects across evaluation districts is \$8,499 (see Figure IV.1). The average maximum pay-for-performance bonus for teachers in tested grades and subjects across non-evaluation districts is \$5,122 (see Figure IV.1). These amounts serve only as a descriptive benchmark because they are based on all evaluation or non-evaluation districts that answered the pay-for-performance amount question, and may not be the same sample of districts that reported additional pay opportunities.

^cSample sizes are presented as a range based on the data available for each row in the table.

*Difference between evaluation and non-evaluation districts is statistically significant at the 0.05 level, two-tailed test.

n.a.= not applicable.

Table D.7. Additional Evidence Used to Evaluate Teacher Performance, as Reported by Educators and District Representatives

Additional Evidence	Percentage of Respondents Reporting That Each Type of Evidence Was Used		
	Teacher Report	Principal Report	District Report
Teacher attendance	41.0	48.9	30.0
Teacher participation in school activities	45.1+	57.1	30.0
Principal's or other administrator's professional judgment	85.2	76.6	70.0
Reviews from other teachers	26.2+	9.6	20.0
Student attendance	8.8+	3.3	10.0
Parent or student input	16.6	13.3	10.0
Sample Size—Range^a	814–817	131–133	10

Source: Teacher, principal, and district surveys.

Notes: Overall values for teacher and principal responses are weighted means so that districts are equally weighted. Overall values for districts are means among the 10 evaluation districts that participated in the educators' survey. Educators' responses are included only if their district responded to the given question. Sample size may vary for individual items due to item nonresponse. The minimum sample size is displayed in the table.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference from the district report is statistically significant at the 0.05 level, two-tailed test.

+Difference between teacher and principal report is statistically significant at the 0.05 level, two-tailed test.

Besides using student achievement growth to evaluate principals—as required by the grant—districts were also required under the core elements to use at least one additional type of evidence. In Table D.8, we compare principals' and districts' reports of whether other types of evidence were used in principal evaluations. Similar percentages of principals and districts reported that assessments by teachers factored into principal evaluations, but principals were more likely than their districts to report that parental input was used (56 percent versus 11 percent).

Table D.8. Additional Evidence Used to Evaluate Principal Performance, as Reported by Principals and District Representatives

Additional Evidence	Percentage of Respondents Reporting That Each Type of Evidence Was Used	
	Principal Report	District Report
Teacher assessments	65.0	60.0
Parent input	55.8*	11.1
Sample Size—Range^a	94–119	9–10

Source: Principal and district surveys.

Note: Overall values for principal responses are weighted means so that districts are equally weighted. Overall values for districts are means among the 10 evaluation districts that participated in the educators' survey. Educators' responses are included only if their district responded to the given question. Sample size may vary for individual items due to item nonresponse. The minimum sample size is displayed in the table.

^aSample 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.

Opportunities for additional pay constituted another component of TIF that grantees had some flexibility to design. Although additional pay for responsibilities was mandated by the grant, districts could have also offered additional pay for other factors, such as teaching in a hard-to-staff school or high-need subject, or participating in professional development. As shown in Table D.9, we found no significant differences among teachers, principals, and districts in their reports of whether teachers were eligible for these other opportunities.

Table D.9. Types of Additional Pay for Teachers, as Reported by Educators and District Representatives

Reason for Additional Pay	Percentage of Respondents Reporting That Teachers Can Receive Additional Pay for the Specified Reason		
	Teacher Report	Principal Report	District Report
Teach in a hard-to-staff school or high-need subject	22.6	29.1	30.0
Teach in a hard-to-staff school	16.9	16.8	20.0
Teach in a high-need subject	15.6	20.2	20.0
Attend professional development activities or enroll in graduate-level courses	33.1	33.6	30.0
Sample Size—Range^a	777–790	132–135	10

Source: Teacher, principal, and district surveys.

Notes: Overall values for teacher and principal responses are weighted means so that districts are equally weighted. Overall values for districts are means among the 10 evaluation districts that participated in the educators’ survey. Educators’ responses are included only if their district responded to the given question. Sample size may vary for individual items due to item nonresponse. The minimum sample size is displayed in the table.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference between teacher or principal and district report is statistically significant at the 0.05 level, two-tailed test.

+Difference between teacher and principal report is statistically significant at the 0.05 level, two-tailed test.

Participation in professional development. In Chapter IV, we reported that most teachers indicated having participated in professional development focused on understanding TIF components. Table D.10 is a summary of these findings. Teachers were asked to report the types of professional development they received in the first half of the school year; principals also provided information about the percentage of their teachers who participated in professional development during this period. In addition, districts reported the percentages of teachers who were expected to receive each type of professional development that would be offered over the course of the full school year. We compared the reports of these different stakeholders to assess the extent to which teachers reported being exposed, during the first half of the year, to the professional development that districts had planned to offer during the full school year. As shown in the table, 71 percent of teachers reported receiving professional development to understand the components of the TIF program. This percentage was lower than that based on principals’ reports (86 percent) and districts’ reports (90 percent).

Table D.10. Percentage of Teachers Receiving Professional Development, as Reported by Educators and District Representatives

	Teacher Report	Principal Report	District Report
Understanding components of TIF program	71.0*+	85.5	89.9
Understanding performance measures of TIF program	67.5*+	80.7	79.0
Feedback based on TIF performance ratings	49.6*+	67.7	53.0
Differentiated instructional strategies based on student assessments	70.6*+	68.1*	24.5
Instructional techniques and strategies	86.1*+	85.7	66.9
Aligning curricula to state or district standards	80.9*+	73.4	56.5
Sample Size—Range^a	802–805	131–133	10–10

Source: Teacher, principal, and district surveys.

Notes: Entries for teacher reports are the percentages of teachers who reported receiving professional development in the specified topic between July 1, 2011 and December 31, 2011. Entries for principal reports are averages of principals' reports about the percentage of teachers in their school who received professional development in the specific topic between July 1, 2011 and December 31, 2011. Both teacher and principal responses are weighted so that districts are equally weighted. Entries for district reports are averages of districts' reports about the percentage of teachers in the district who were expected to receive the professional development being planned in the specific topic for the full-year period from July 1, 2011 to June 30, 2012. Educators' responses are included only if their district responded to the given question. Sample size may vary for individual items due to item nonresponse. Table displays the minimum sample size.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Difference between teacher or principal and district report is statistically significant at the 0.05 level, two-tailed test.

+Difference between teacher and principal report is statistically significant at the 0.05 level, two-tailed test.

Subgroup analyses. In Chapter IV, we indicated that a larger percentage of teachers in tested grades and subjects than in nontested grades and subjects reported being evaluated on measures of student achievement growth, and the details of these findings are in Table D.11. Across all levels at which student achievement growth could be aggregated (schools, student groups, and classrooms), teachers in tested grades and subjects were 11 to 13 percentage points more likely than those in nontested grades and subjects to report being evaluated on the growth measure. As discussed in Chapter IV, there were no significant differences between these subgroups in their reported likelihood of being evaluated by classroom observations.

Table D.11. Teacher Performance Measures Used, as Reported by Teachers in Tested and Nontested Grades and Subjects

	Percentage of Teachers Reporting That the Measure Was Used		
	Tested Grades and Subjects	Nontested Grades and Subjects	Difference
Achievement Level	65.3	50.8	14.5*
Achievement Growth	75.2	64.1	11.1*
By school	68.5	55.5	13.0*
By student group ^a	61.4	49.7	11.7*
By teacher's classroom	66.2	54.8	11.4*
Classroom Observations	80.1	78.4	1.7
Number of Teachers—Range^b	482–485	325–328	

Source: Teacher survey.

^aStudent groups include grade levels, teams, and subject areas.

^bSample 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.

Teachers' reports of pay-for-performance eligibility also had the potential to vary across subgroups. In particular, because three districts were known to have offered pay-for-performance to control schools (contrary to the intentions of the grant), there was the potential for the treatment-control difference in reported pay-for-performance eligibility to be smaller in those districts than in the remaining districts. One of these districts offered a bonus of up to \$800 to teachers who earned a high rating on the district's teacher evaluation. The other two districts offered bonuses of up to \$2,000 based on school performance. However, according to the grantee, teachers expected to receive the bonus each year because nearly all schools earned it. However, there was no evidence, as indicated in Table D.12, of smaller treatment-control differences in the three districts that offered pay-for-performance to control schools. That table also shows teachers' reports of pay-for-performance eligibility within the other types of subgroups in which we estimated intermediate impacts, as reported in Chapter V and Appendix E.

Exclusion of imputed maximum bonus amounts. As discussed in Appendix C, we used multiple imputation to address item nonresponse by teachers and principals on survey questions that asked about the largest possible pay-for-performance bonus for which they were eligible. In Chapter IV (Figures IV.8 and IV.12), we presented teachers' and principals' average report of the largest possible pay-for-performance bonus, including these imputed values. The corresponding findings without the imputed values are shown in Figures D.2 and D.3.

Table D.12. Teachers' Eligibility for Pay-for-Performance Bonuses, as Reported by Teachers: Subgroup Analyses

Subgroup	Percentage of Teachers Reporting That Teachers in Their Schools Were Eligible for Pay-for-Performance Bonuses			Number of Teachers
	Treatment	Control	Difference	
All Teachers (primary analysis)	48.0	17.3	30.7*	787
Teaching Assignment				
(1) Tested grades and subjects	46.5	14.8	31.7*	466
(2) Nontested grades and subjects	50.2	20.7	29.5*	321
Difference between subgroups, (1) - (2)			2.2	
Teacher Experience				
(1) Less than 5 years	54.4	15.7	38.7*	241
(2) 5 to 24	45.9	19.3	26.6*	467
(3) Greater than 24	41.5	10.6	30.9*	79
Difference between subgroups, (1) - (2)			12.0	
Difference between subgroups, (3) - (2)			4.2	
Type of Approach to Teacher Evaluation ^a				
(1) No teacher-level growth	43.1	9.0	34.1*	308
(2) Emphasize teacher-level growth	44.5	24.4	20.0*	361
(3) Combine teacher and school growth (TAP)	75.1	23.5	51.6*	118
Difference between subgroups, (1) - (2)			14.0*	
Difference between subgroups, (3) - (2)			31.6*	
District Maximum Pay-for-Performance Bonus Amount ^b				
(1) High (above median)	49.0	16.9	32.1*	528
(2) Low (below median)	46.1	18.1	28.0*	259
Difference between subgroups, (1) - (2)			4.1	
District Offering of Pay-for-Performance Bonus to Controls ^c				
(1) Offered pay-for-performance bonus	43.6	20.2	23.3*	162
(2) Did not offer pay-for-performance bonus	49.2	16.5	32.8*	625
Difference between subgroups (1) - (2)			-9.4	

Source: Teacher survey, district survey, technical assistance documents, and district interviews.

Notes: The difference between the treatment and control group is adjusted for block fixed effects. The mean outcome for the treatment group is calculated as the unadjusted mean outcome for control group plus the adjusted difference in outcomes between the two groups. The primary model for all teachers estimates a weighted average of the district-specific impacts estimates as described in Appendix D. Subgroups means and hypothesis testing are based on a model with a treatment dummy and interaction(s) between the treatment and the subgroup(s) using the pooled sample.

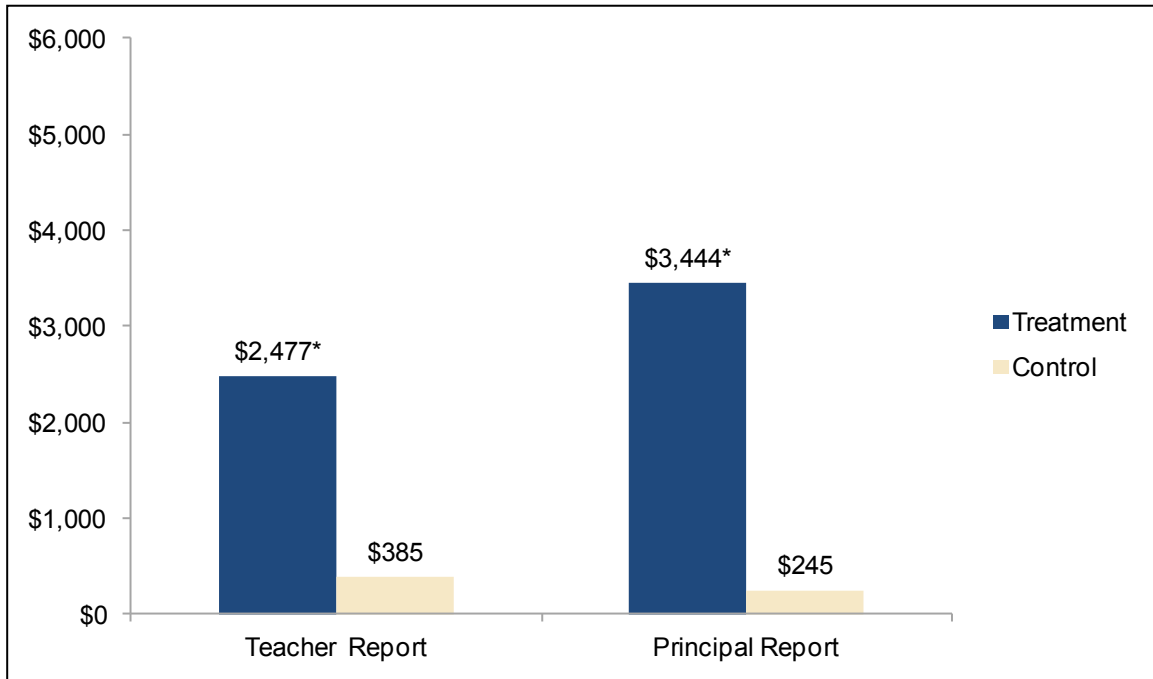
^aTypology is based on technical assistance documents.

^bPay-for-performance bonus amount is calculated based on a combination of survey questions and district interviews as described in Appendix D.

^cInformation on the offering of pay-for-performance bonuses to control schools was obtained through technical assistance.

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

Figure D.2. Maximum Possible Size of Pay-for-Performance Bonuses for Teachers, as Reported by Teachers and Principals Who Provided Nonmissing Responses

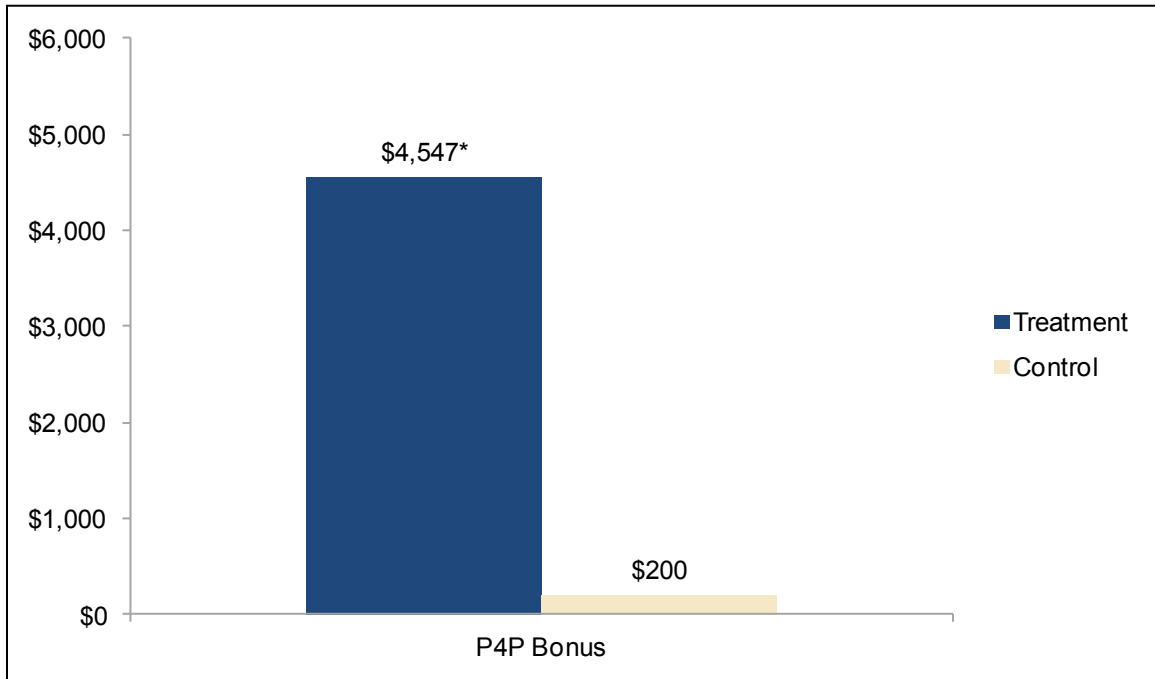


Source: Teacher and principal surveys.

Notes: Figures indicate respondents' average report of the maximum possible size of teachers' pay-for-performance bonuses. A total of 348 treatment teachers, 378 control teachers, 53 treatment principals, and 58 control principals responded to this survey question. No missing values were imputed.

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

Figure D.3. Maximum Possible Size of Pay-for-Performance Bonuses for Principals, as Reported by Principals Who Provided Nonmissing Responses



Source: Principal survey.

Notes: Figures indicate respondents' average report of the maximum possible size of their pay-for-performance bonuses. A total of 59 treatment principals and 62 control principals responded to the survey question on pay-for-performance bonuses. No missing values were imputed.

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

APPENDIX E
SUPPLEMENTARY FINDINGS FOR CHAPTER V

THIS PAGE IS INTENTIONALLY BLANK

This appendix supplements Chapter V findings on the impacts of pay-for-performance in three ways: (1) additional analyses to help capture the intensity of attitudes and beliefs, (2) sensitivity analyses to examine the robustness of the impact findings to a key modeling assumption, and (3) a subgroup analysis to shed light on possible differential impacts of pay-for-performance.

Intensity of Teacher and Principal Attitudes

First, we present additional analyses to assess impacts on the intensity of attitudes and beliefs. In Chapter V, we examined outcomes for which we collapsed four-category Likert scales in most cases into dichotomous measures (“agree” or “strongly agree” versus “disagree” or “strongly disagree”). This rule makes the important qualitative distinctions between satisfaction and dissatisfaction or between agreement and disagreement but does not capture intensity of feeling. In Tables E.1 through E.5, we present the impacts based on constructs presented in Chapter V alongside the impact estimates for the alternative construct, which is based on the percentage of respondents reporting the “top” category only (“very satisfied” or “strongly agree”) versus all other responses.

Table E.1. Impacts on Satisfaction Using Alternative-Outcome Definitions

Respondent Type and Satisfaction Measure	Treatment-Control Difference in Percent Reporting	
	“Somewhat” or “Very” Satisfied	“Very” Satisfied
Teachers		
Use of Measures of Performance		
Classroom observations	-8.6*	-2.4
Student achievement	-2.0	-4.8*
Opportunities for Pay and Development		
Opportunities for professional advancement	-7.8*	-3.7
Opportunities to enhance skills	-1.3	-3.7
Opportunities to earn extra pay	5.1*	1.9
School Environment		
Recognition of accomplishments	-5.4	-5.1
Quality of interaction with colleagues	-7.0*	-8.5*
Colleagues’ efforts	-1.6	-5.8*
School morale	-6.8*	-8.7*
Job Satisfaction		
Overall job satisfaction	-5.3	-4.0
Number of Teachers—Range^a	810–820	810–820
Principals		
Opportunities for Pay and Development		
Opportunities to enhance skills	-2.2	3.2
Opportunities to earn extra pay	5.1	21.4*
Intellectual challenge	0.5	10.0
Feedback on Performance		
School environment	-4.2	7.6
Recognition of accomplishments	-4.8	5.9
Quality of interaction with colleagues	-6.9	10.2
Colleagues’ efforts	-5.8	2.2
Colleagues’ contribution to student learning	-6.2*	-4.6
School morale	-16.4*	4.0
Number of Principals—Range^a	128–130	128–130

Source: Teacher and principal surveys.

^aSample 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 E.2. Teachers' Attitudes Toward TIF Program Using Alternative-Outcome Specifications

Statement	Treatment-Control Difference in Percent Reporting	
	"Agree" or "Strongly Agree"	"Strongly Agree"
Teachers who do the same job should receive the same pay	-0.5	-1.3
Standardized student test scores in my district measure what students have learned	1.1	0.1
My principal is a good judge of teacher talent	-7.1*	-4.4*
I am glad that I am participating in the TIF program	2.1	-0.2
My job satisfaction has increased due to the TIF program	-4.9*	-0.7
I feel increased pressure to perform due to the TIF program	8.7*	7.9*
I have less freedom to teach the way I would like to teach due to the TIF program	1.1	2.3
The TIF program has harmed the collaborative nature of teaching	1.3	-0.1
The TIF program has caused teachers to work more effectively	3.8	3.7*
The TIF program is fair	-4.6*	-2.3*
The process used to determine how bonuses are determined was adequately explained to me	7.8*	0.6
Number of Teachers—Range^a	793–815	793–815

Source: Teacher survey.

^aSample 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 E.3. Principals' Attitudes Toward TIF Program Using Alternative-Outcome Specifications

Statement	Treatment-Control Difference in Percent Reporting	
	"Agree" or "Strongly Agree"	"Strongly Agree"
The TIF program has been clearly communicated to me	-7.0	-2.6
This school has less chance of earning a bonus because of the characteristics of our student population	4.9	3.3
The evaluation system omits important aspects of school administration that should be considered	1.2	3.4
The TIF program contributes to greater collegiality and professionalism among the staff at this school	-7.0	3.1
Teachers at this school are more comfortable with frequent formal observations of their teaching because of the TIF program	-5.4	3.3
Parents and the school community believe the TIF program is important	-7.2	1.8
The TIF program is likely to continue for the foreseeable future	-4.9	-5.0
I played an important role in implementing the TIF program at my school	-1.1	6.5
Number of Principals—Range^a	129–135	129–135

Source: Principal survey.

Note: Overall values are unadjusted means.

^aSample 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 E.4. Incentives Used to Recruit Teachers Using Alternative-Outcome Specifications

Incentives Used for Recruiting Teachers	Treatment-Control Difference in Percent Reporting	
	"Always" or "Often" Used	"Always" Used
Salary	0.3	-2.9
Opportunities to earn pay-for-performance	9.8*	1.8
Opportunities for career advancement	2.7	6.8*
Opportunities for professional development	-0.6	-0.5
The level of teacher involvement in school decision making	-6.1	4.3
Collegiality of teaching staff	-8.3	5.1
The school culture and/or educational philosophy	-2.6	-2.9
The school's reputation	1.5	2.0
The school's location or neighborhood	-0.3	10.0*
The level of student achievement at the school	0.6	10.6*
The TIF program	16.4*	11.6*
Number of Principals—Range^a	129–134	129–134

Source: Principal survey.

^aSample 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 E.5. Criteria Used for Teacher Assignments to Grade Levels or Subject Areas Using Alternative-Outcome Specifications

Criterion	Treatment-Control Difference in Percent Reporting	
	"Always" or "Often" Used	"Always" Used
The teacher's experience in a grade level or subject area	1.4	-15.2*
The teacher's seniority	-9.9*	1.0
The teacher's content knowledge	-4.3	-13.7
The teacher's ability to produce high test scores in grades/classes in which state or federal assessments are administered	-0.1	-10.2*
The teacher's ability to work with certain student populations	6.2	-2.7
To balance teacher experience and expertise in a grade level or subject	2.3	-10.0
Number of Principals—Range^a	132–135	132–135

Source: Principal survey.

^aSample 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.

Robustness to Assumptions on Blocked Fixed Effects and Linearity

In Chapter V, we used a linear regression to estimate impacts, but here we examine the robustness of those findings to the use of an alternative approach. We called the linear regression the primary model. In Tables E.6 through E.11, we present the primary results alongside two alternative models. Tables E.6 and E.7 correspond to the Chapter V findings on attitudes. Tables E.8 through E.10 correspond to the Chapter V findings on self-reported behaviors. Table E.11 provides information on teachers' characteristics, which are presented in detail later in this appendix. Those tables are used to illustrate the possible early effect pay-for-performance may have had on the composition of the teacher workforce.

The difference between the models lies in how we treat the outcomes, which are binary (equal to zero or 1). The relationship between covariates and binary outcomes is approximately linear, especially when the proportion of the sample exhibiting a positive outcome is near 50 percent, but is nonlinear as the percentages approach 0 or 100. We used the linear model as the primary analysis because it is easier to estimate and understand. The logit model is a more appropriate specification for binary outcomes, but it can be unstable in the presence of fixed effects, which we rely upon in this case to account for the study design. Therefore, we estimated the logit models with block-treatment weights (such that each combination of treatment status and block had the same weight) instead of block fixed effects.

To make fair comparisons between linear and logit models, we estimated a second version of the linear regression model using the same block-treatment weights and no block fixed effects. We compared all three models, referred to as A, B, and C, in the tables below. Models A and B differ from each other only in the use of an overall treatment effect and whether they include block fixed effects. Models B and C differ from each other only in whether they use the linear or logit regression.

Table E.6. Teacher and Principal Satisfaction with Professional Opportunities and School Environment Using Alternative-Model Specifications (Percentages Who Are “Somewhat” or “Very” Satisfied)

Model	Treatment-Control Difference		
	Linear Regression (Primary Analysis)	Linear Regression	Logit Marginal Effects
	A	B	C
<i>Weighted Average of District-Specific Impacts</i>	✓		
<i>One Overall Treatment Effect</i>		✓	✓
<i>Block Fixed Effects</i>	✓		
<i>Block-Treatment Weights</i>		✓	✓
Teachers			
Use of Measures of Performance			
Classroom observations	-8.6*	-8.6	-8.5
Student achievement	-2.0	-0.9	-0.9
Opportunities for Pay and Development			
Opportunities for professional advancement	-7.8*	-6.2	-6.2
Opportunities to enhance skills	-1.3	-2.3	-2.3
Opportunities to earn extra pay	5.1*	4.2	4.2
School Environment			
Recognition of accomplishments	-5.4	-3.1	-3.1
Quality of interaction with colleagues	-7.0*	-6.7	-6.7
Colleagues' efforts	-1.6	-1.6	-1.6
School morale	-6.8*	-6.8	-6.7
Job Satisfaction			
Overall job satisfaction	-5.3	-4.9	-4.9
Number of Teachers—Range^b	810–820	810–820	810–820
Principals			
Opportunities for Pay and Development			
Opportunities to enhance skills	-2.2	-2.6	-2.6
Opportunities to earn extra pay	5.1	5.1	5.1
Intellectual challenge	0.5	-3.0	-3.1
Feedback on Performance	-4.2	-4.6	-4.6
School Environment			
Recognition of accomplishments	-4.8	-9.6	-9.6
Quality of interaction with colleagues	-6.9	-11.5*	-13.6
Colleagues' efforts	-5.8	-9.7	-11.4
Colleagues' contribution to student learning	-6.2*	-8.9*	^a
School morale	-16.4*	-17.0*	-17.2*
Number of Principals—Range^b	128–130	128–130	128–130

Source: Teacher and principal surveys.

Notes: Model A: The primary model (used in the main body of the report) is a linear probability model with randomization block fixed effects. The estimated impact is the weighted average of the district-specific impacts estimates, with each district weighted by the number of schools in the evaluation. Teachers are weighted such that each school contributes equally to the average-impact estimate. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model B: The linear regression model is a linear probability model using weights so that each block-treatment combination is weighted equally. The model estimates one overall impact across all districts. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model C: The logit regression model follows the same model specification as the linear regression with block-treatment weights, but estimates it using a logit instead of a linear regression. Marginal effects are shown in the table.

^aImpact was not estimable due to the lack of variation in the outcome within one of the treatment status groups.

^bSample 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 E.7. Teachers' Attitudes Toward TIF Program Using Alternative-Model Specifications (Percentages Who "Agree" or "Strongly Agree")

Model	Treatment-Control Difference		
	Linear Regression (Primary Analysis)	Linear Regression	Logit Marginal Effects
	A	B	C
<i>Weighted Average District Impacts</i>	✓		
<i>One Overall Treatment Effect</i>		✓	✓
<i>Block Fixed Effects</i>	✓		
<i>Block-Treatment Weights</i>		✓	✓
Statement			
Teachers who do the same job should receive the same pay	-0.5	0.7	0.7
Standardized student test scores in my district measure what students have learned	1.1	1.3	1.3
My principal is a good judge of teacher talent	-7.1*	-7.5	-7.5
I am glad that I am participating in the TIF program	2.1	4.0	4.0
My job satisfaction has increased due to the TIF program	-4.9*	-5.2	-5.2
I feel increased pressure to perform due to the TIF program	8.7*	4.7	4.7
I have less freedom to teach the way I would like to teach due to the TIF program	1.1	1.8	1.8
The TIF program has harmed the collaborative nature of teaching	1.3	1.8	1.8
The TIF program has caused teachers to work more effectively	3.8	0.6	0.6
The TIF program is fair	-4.6*	-3.9	-3.9
The process used to determine how bonuses are determined was adequately explained to me	7.8*	6.4	6.3
Number of Teachers—Range^a	793–815	793–815	793–815

Source: Teacher survey.

Notes: Model A: The primary model (used in the main body of the report) is a linear probability model with randomization block fixed effects. The estimated impact is the weighted average of the estimates of the district-specific impacts, with each district weighted by the number of schools in the evaluation. Teachers are weighted such that each school contributes equally to the average-impact estimate. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model B: The linear regression model is a linear probability model using weights so that each block-treatment combination is weighted equally. The model estimates one overall impact across all districts. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model C: The logit regression model follows the same model specification as the linear regression with block-treatment weights, but estimates it using a logit instead of a linear regression. Marginal effects are shown.

^aSample 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 E.8. Incentives Used to Recruit Teachers Using Alternative-Model Specifications (Percentages Who Report They Are “Always” or “Often” Used)

Model	Treatment-Control Difference		
	Linear Regression (Primary Analysis)	Linear Regression	Logit Marginal Effects
	A	B	C
<i>Weighted Average District Impacts</i>	✓		
<i>One Overall Treatment Effect</i>		✓	✓
<i>Block Fixed Effects</i>	✓		
<i>Block-Treatment Weights</i>		✓	✓
Incentives Used for Recruiting Teachers			
Salary	0.3	1.0	1.0
Opportunities to earn pay-for-performance	9.8*	4.5	4.5
Opportunities for career advancement	2.7	-2.4	-2.4
Opportunities for professional development	-0.6	-0.5	-0.5
The level of teacher involvement in school decision making	-6.1	-10.9	-10.8
Collegiality of teaching staff	-8.3	-9.8	-9.8
The school culture and/or educational philosophy	-2.6	-3.4	-3.4
The school’s reputation	1.5	3.2	3.2
The school’s location or neighborhood	-0.3	1.0	1.0
The level of student achievement at the school	0.6	-2.1	-2.1
The TIF program	16.4*	15.4	15.2
Number of Principals—Range^a	129–134	129–134	129–134

Source: Principal survey.

Notes: Model A: The primary model (used in the main body of the report) is a linear probability model with randomization block fixed effects. The estimated impact is the weighted average of the estimates of the district-specific impacts, with each district weighted by the number of schools in the evaluation. Teachers are weighted such that each school contributes equally to the average-impact estimate. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model B: The linear regression model is a linear probability model using weights so that each block-treatment combination is weighted equally. The model estimates one overall impact across all districts. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model C: The logit regression model follows the same model specification as the linear regression with block-treatment weights, but estimates it using a logit instead of a linear regression. Marginal effects are shown in the table.

^aSample 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 E.9. Criteria Principals Used for Assigning Teachers to Grade Levels or Subject Areas Using Alternative-Model Specifications (Percentages Who Report They Are “Always” or “Often” Used)

Model	Treatment-Control Difference		
	Linear Regression (Primary Analysis)	Linear Regression	Logit Marginal Effects
	A	B	C
<i>Weighted Average District-impacts</i>	✓		
<i>One Overall Treatment Effect</i>		✓	✓
<i>Block Fixed Effects</i>	✓		
<i>Block-Treatment Weights</i>		✓	✓
Criteria			
The teacher’s experience in a grade level or subject area	1.4	-1.9	-1.9
The teacher’s seniority	-9.9*	-9.8	-10.0
The teacher’s content knowledge	-4.3	-0.7	-0.7
The teacher’s ability to produce high test scores in grades/classes in which state or federal assessments are administered	-0.1	1.3	1.3
The teacher’s ability to work with certain student populations	6.2	8.5	8.5
To balance teacher experience and expertise in a grade level or subject	2.3	6.0	6.0
Number of Principals—Range^a	132–135	132–135	132–135

Source: Principal survey.

Notes: Model A: The primary model (used in the main body of the report) is a linear probability model with randomization block fixed effects. The estimated impact is the weighted average of the estimates of district-specific impacts, with each district weighted by the number of schools in the evaluation. Teachers are weighted such that each school contributes equally to the average-impact estimate. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model B: The linear regression model is a linear probability model using weights so that each block-treatment combination is weighted equally. The model estimates one overall impact across all districts. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model C: The logit regression model follows the same model specification as the linear regression with block-treatment weights, but estimates it using a logit instead of a linear regression. Marginal effects are shown in the table.

^aSample 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 E.10. Influence of TIF Program on Educators' School Preference (Percentages)

Model	Treatment-Control Difference		
	Linear Regression (Primary Analysis)	Linear Regression	Logit Marginal Effects
	A	B	C
<i>Weighted Average District-impacts</i>	✓		
<i>One Overall Treatment Effect</i>		✓	✓
<i>Block Fixed Effects</i>	✓		
<i>Block-Treatment Weights</i>		✓	✓
Teachers			
TIF program affected where or what to teach	1.9*	1.7	1.7
Ways in which TIF affected where or what to teach			
Stayed at school because of TIF	1.3*	0.8	0.9
Changed school to get into TIF	-0.3	0.2	0.2
Changed primary grade or subject because of TIF	-0.3	0.3	0.3
Applied to current school to get into TIF	1.1*	0.7	0.7
Applied for position in another school to leave TIF	-0.2	-0.2	--
Applied for position in another school with better bonus program	0.0	0.0	--
TIF program expected to affect preference of school for next year	2.2	2.1	2.1
Ways in which TIF program will affect school preference			
Stay at current school because of TIF	2.0	3.0	3.0
Change school to get out of TIF	-2.1*	-2.5	-2.7
Change grade or subject because of TIF	0.2	-0.3	-0.3
Apply for position in another school to leave TIF	1.0	0.9	0.9
Apply for position in another school with better bonus program	0.8	0.9	0.9
Number of Teachers—Range^a	824–825	824–825	824–825
Principals			
TIF program affected choice of school	3.8	4.9	5.0
Ways in which TIF affected school preference			
Stayed at school because of TIF	6.7*	5.8	7.0
Came to school to get into TIF	-2.9	-0.9	-0.9
Number of Principals	134	134	134

Source: Teacher and principal surveys.

Notes: Model A: The primary model (used in the main body of the report) is a linear probability model with randomization block fixed effects. The estimated impact is the weighted average of the estimates of district-specific impacts, with each district weighted by the number of schools in the evaluation. Teachers are weighted such that each school contributes equally to the average-impact estimate. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model B: The linear regression model is a linear probability model using weights so that each block-treatment combination is weighted equally. The model estimates one overall impact across all districts. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model C: The logit regression model follows the same model specification as the linear regression with block-treatment weights, but estimates it using a logit instead of a linear regression. Marginal effects are shown in the table.

^aSample 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 E.11. Demographic Characteristics, Educational Background, and Certification of Teachers (Percentages)

Model	Treatment-Control Difference		
	Linear Regression (Primary Analysis)	Linear Regression	Logit Marginal Effects
	A	B	C
<i>Weighted Average District-impacts</i>	✓		
<i>One Overall Treatment Effect</i>		✓	✓
<i>Block Fixed Effects</i>	✓		
<i>Block-Treatment Weights</i>		✓	✓
Characteristic			
Female	4.1*	4.4	4.5
Race/Ethnicity			
White, non-Hispanic	5.3*	1.7	1.7
Black, non-Hispanic	-5.8*	-2.5	-2.5
Hispanic	0.4	-0.2	-0.2
Other	0.0	1.0	1.0
Married	-1.5	-5.6	-5.5
Children under 18 years in Household	1.2	0.9	0.9
Number of Teachers—Range^a	803–820	803–820	803–820
Background			
Master's Degree or Higher	-4.4	-5.5	-5.5
Bachelor's Degree			
From a highly selective or selective college or university ^a	1.8	-1.4	-1.4
Major			
Elementary education	7.6*	8.3	8.3
Secondary education	0.3	0.7	0.7
Other education	-0.4	0.3	0.3
Subject matter-specific	-7.5*	-9.3	-9.2
Certification			
Certification Status			
Regular or standard certificate	4.9*	3.0	3.0
Certified for Current Teaching Position ^b	-1.2	-1.5	-1.5
Certification Route			
Traditionally certified	0.5	-3.5	-3.6
National Board Certified	-0.7	-1.2	-1.2
Number of Teachers—Range^a	700–826	700–826	700–826

Source: Teacher survey.

Notes: Model A: The primary model (used in the main body of the report) is a linear probability model with randomization block fixed effects. The estimated impact is the weighted average of the estimates of district-specific impacts, with each district weighted by the number of schools in the evaluation. Teachers are weighted such that each school contributes equally to the average-impact estimate. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model B: The linear regression model is a linear probability model using weights so that each block-treatment combination is weighted equally. The model estimates one overall impact across all districts. Standard errors are robust and clustered at the unit of random assignment (either schools or groups of schools).

Model C: The logit regression model follows the same model specification as the linear regression with block-treatment weights, but estimates it using a logit instead of a linear regression. Marginal effects are shown in the table.

^aSample 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.

Impacts on Teacher Attitudes by Subgroup

Next, we present findings by subgroup in Tables E.12 to E.15. All of the major teacher and principal findings reported in Chapter V appear in these tables, but with separate groupings of teachers or districts based on teaching assignment (tested or nontested grade-subjects), teacher experience (early-, mid-, or late-career), or district program type and generosity, as measured by the size of the maximum pay-for-performance payout. Each column is an outcome and each row is a separate subgroup, except for the last two tables, where results are presented separately by grade span or by the respondent's answer to a prior screening question.

Table E.12. Teacher Satisfaction by Subgroup (Percentages That Are “Somewhat” or “Very” Satisfied)

Subgroup	Use of Classroom Observations To Measure 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	Efforts of My Colleagues	School Morale	Overall Job Satisfaction	Number of Teachers
Treatment-Control Difference											
All Teachers (primary analysis)	-8.6*	-2.0	-7.8*	-1.3	5.1*	-5.4	-7.0*	-1.6	-6.8*	-5.3	810
Teaching Assignment											
(1) Tested grades and subjects	-8.9*	-1.2	-8.8*	-3.5	2.4	-9.1	-10.4*	-3.6	-7.3	-8.6	485
(2) Nontested grades and subjects	-7.7	-3.0	-6.3	2.2	8.9	0.1	-1.6	1.7	-5.5	-0.1	325
Difference between subgroup (1) - (2)	-1.2	1.7	-2.4	-5.8	-6.5	-9.2	-8.8	-5.3	-1.8	-8.5	
Teacher Experience											
(1) Less than 5 years	-3.7	12.4	-13.8*	-7.2	7.9	2.3	3.1	5.1	11.4	3.8	250
(2) 5 to 24 years	-8.8*	-6.2	-5.7	3.9	6.5	-6.2	-6.5*	-2.4	-10.7*	-7.1	482
(3) Greater than 24 years	-20.4	-18.8	-4.3	-14.9	-13.7	-23.0*	-38.3*	-15.7	-36.0*	-21.1*	77
Difference between subgroups (1) - (2)	5.1	18.6	-8.1	-11.1	1.4	8.5	9.6	7.5	22.1	10.9	
Difference between subgroups (3) - (2)	-11.6	-12.6	1.4	-18.8	-20.2	-16.8	-31.8*	-13.2	-25.3	-14.1	
District Program Type ^a											
(1) No Teacher-Level Growth	-13.3*	-0.7	-7.7*	-0.5	4.2	-10.9*	-10.3*	-4.7	-17.1*	-12.0*	314
(2) Emphasize Teacher-Level Growth	-7.3	-6.5	-11.5*	-3.8	0.4	-4.2	-5.4*	2.0	-4.8	-2.5	374
(3) Combine Teacher and School Growth (TAP)	2.9	7.4	2.4	3.9	22.0*	8.7	-0.6	-1.9	20.4*	7.9	121

E.17

Table E.12 (continued)

Subgroup	Use of Classroom Observations To Measure 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	Efforts of My Colleagues	School Morale	Overall Job Satisfaction	Number of Teachers
Difference between subgroups (1) - (2)	-6.0*	5.8*	3.8*	3.3*	3.7*	-6.7*	-4.9*	-6.7	-12.3	-9.5*	
Difference between subgroups (3) - (2)	10.1	14.0	13.9	7.7	21.6*	13.0	4.8*	-3.8*	25.2*	10.4	
District Maximum Pay-for-Performance Bonus Amount ^b											
(1) High (above median)	-11.6*	-4.6	-9.6*	-2.8	3.0	-7.1	-10.4*	-3.4	-9.3*	-6.2	543
(2) Low (below median)	-2.9	3.0	-4.5	1.6	9.1	-2.3	-0.3	1.9	-2.0	-3.5	267
Difference between subgroups (1) - (2)	-8.7	-7.6	-5.1	-4.4	-6.1	-4.8	-10.1*	-5.3	-7.3	-2.7	

Source: Teacher survey, district survey, technical assistance documents, and district interviews.

Note: The difference between treatment and control group is adjusted for block fixed effects. The primary model for all teachers estimates a weighted average of the district-specific impacts as described in Appendix D. 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) using the pooled sample.

^aProgram Type is a typology based on technical assistance documents.

^bPay-for-performance bonus amount is calculated based on a combination of district survey questions and district interviews, as described in Appendix C.

*Difference is statistically significant at the 0.05 level, two-tailed test.

Table E.13. Teachers' Attitudes Toward TIF Program by Subgroup (Percentages Who "Agree" or "Strongly Agree")

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
Treatment-Control Difference												
All Teachers (primary analysis)	-0.5	1.1	-7.1*	2.1	-4.9*	8.7*	1.1	1.3	3.8	-4.6*	7.8*	793
Teaching Assignment												
(1) Tested grades and subjects	4.8	2.4	-7.7	2.1	-5.4	8.4	2.4	2.9	3.5	-6.1	7.3	471
(2) Nontested grades and subjects	-7.7	-0.5	-5.6	2.1	-4.2	8.8	-0.7	-1.3	4.4	-2.6	8.7	320
Difference between subgroup (1) - (2)	12.5	2.9	-2.1	0.0	-1.2	-0.4	3.1	4.2	-0.9	-3.5	-1.3	
Teacher Experience												
(1) Less than 5 years	-4.4	-1.7	5.1	3.9	-4.7	16.4*	2.4	-3.3	16.7*	-8.2	20.0*	243
(2) 5 to 24 years	2.5	6.0	-10.6*	4.7	-1.7	7.4	-2.8	0.7	2.9	-0.4	3.5	474
(3) Greater than 24 years	-7.7	-19.8	-21.1	-17.6	-25.2*	-6.9	18.4	17.6	-28.2*	-18.5	-4.9	75
Difference between subgroups (1) - (2)	-6.9	-7.7	15.7	-0.8	-2.9	8.9	5.2	-4.0	13.9	-7.7	16.5*	
Difference between subgroups (3) - (2)	-10.2	-25.8	-10.5	-22.4	-23.5	-14.3	21.3	16.9	-31.0	-18.1	-8.4	
District Program Type ^a												
(1) No Teacher-Level Growth	1.1	1.6	-16.6*	-2.2	-7.3*	10.1*	6.0	7.0*	-1.4	-9.6*	0.4	311
(2) Emphasize Teacher-Level Growth	-5.4	-4.4	0.7	2.6	-9.1*	7.3	0.1	1.6	3.6	-5.0	10.4*	371
(3) Combine Teacher and School Growth (TAP)	8.9	15.7*	0.2	13.7	14.8	8.3	-11.4	-18.1*	20.5	12.0	23.4*	109

E.19

Table E.13 (continued)

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
Difference between subgroups (1) - (2)	6.5*	5.9*	-17.3*	-4.8*	1.8*	2.8*	5.9*	5.4*	-5.0	-4.7*	-10.1*	
Difference between subgroups (3) - (2)	14.2	20.1*	-0.5	11.2	23.8	1.0*	-11.5*	-19.7	16.9*	17.0	13.0*	
District Maximum Pay-for-Performance Bonus Amount ^b												
(1) High (above median)	-3.9	-2.6	-7.4	4.9	-4.6	11.4*	1.4	1.2	8.7*	-3.6	8.9*	543
(2) Low (below median)	6.0	8.1	-6.4	-3.5	-5.7	3.4	0.5	1.4	-5.8	-6.8	5.5	250
Difference between subgroups (1) - (2)	-9.9*	-10.7	-1.0	8.3	1.1	8.1	1.0	-0.2	14.6	3.2	3.5	

Source: Teacher survey, district survey, technical assistance documents, and district interviews.

Note: The difference between treatment and control group is adjusted for block fixed effects. The primary model for all teachers estimates a weighted average of the district-specific impacts as described in Appendix D. 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) using the pooled sample.

^aProgram Type is a typology based on technical assistance documents.

^bPay-for-performance bonus amount is calculated based on a combination of district survey questions and district interviews, as described in Appendix C.

*Difference is statistically significant at the 0.05 level, two-tailed test.

Teacher Characteristics

We used the background survey to tabulate the characteristics of treatment and control teachers as of the time of the survey (spring 2012). This provides a description of the sample, but also, by comparing the treatment and control groups, an estimate of the early impact of pay-for-performance on the composition of the teacher workforce.

Teachers' decisions about whether to return to the school, or, in some cases, whether to leave during the school year, could have been influenced by the presence or absence of pay-for-performance.⁴⁹ Evidence presented in Chapter V suggests that principals made the pay-for-performance differences a part of their recruitment strategy. Also, there was enough lead time for such effects. Over one-third (36 percent) of schools were randomly assigned before April 2011, early enough to precede the usual teacher transfer process for 2011–2012 in most districts, and the rest were assigned through the end of June 2011, when the school year would have ended.

If there were no recruitment or retention effects, then random assignment will have produced teachers with similar characteristics, on average, in treatment and control schools. Any systematic differences, on the other hand, would be the result of transfers in or out that were induced by treatment. We noted above that there was already a difference reported by principals in the number of vacancies to fill, which suggested that retention in treatment schools was greater. The average difference was 1.5 vacancies per school.

Teachers' background characteristics, summarized in Tables E.14 through E.17, suggest that pay-for-performance eligibility did have an impact on teacher mobility. There were statistically significant treatment-control differences in teachers' demographic characteristics (Table E.14) and professional characteristics (Tables E.15 and E.16). Treatment teachers were 4 percentage points more likely to be female and 6 percentage points less likely to be African American. There were no statistically significant treatment-control differences in average age, percentage of teachers who were married, or percentage who had children under 18 years of age.

Although there were no significant differences in the percentage of teachers with a master's degree or higher, or the selectivity of the undergraduate college attended, treatment teachers held different college degree majors (60 percent versus 52 percent in elementary; 35 percent and 42 percent in subject matter), an effect driven by elementary teachers (see Table E.17).⁵⁰ In addition, treatment teachers were more likely (by 5 percentage points) to have regular or standard certification. There was, however, no difference in in-field teaching rate, certification route, or National Board certification rate.

⁴⁹ Supporting this notion that teacher moves could have happened after random assignment, as intended by the study design (Glazer et al. 2011), is reinforced by our finding that 82 percent of teachers who were new to their schools reported accepting their position on a date that was later than the random assignment date for their school.

⁵⁰ Because random assignment was stratified by grade level, and survey respondents by teaching assignment are similar (Appendix B, Table B.3), treatment and control group differences cannot be explained by the estimation sample.

Table E.14. Demographic Characteristics of Teachers and Principals (Percentages Unless Otherwise Noted)

	Teachers			Principals		
	Treatment	Control	Impact	Treatment	Control	Impact
Female	89.1	85.0	4.1*	60.0	65.6	-5.7
Race/Ethnicity						
White, non-Hispanic	72.7	67.5	5.3*	59.2	54.8	4.3
Black, non-Hispanic	17.6	23.3	-5.8*	30.2	38.7	-8.5
Hispanic	5.6	5.1	0.4	6.0	4.8	1.2
Other	4.1	4.1	0.0	4.6	1.6	3.0
Age (average years)	39.9	39.8	0.1	47.6	48.0	-0.4
Married	66.2	67.7	-1.5	n.a.	n.a.	
Children under 18 years in Household	46.7	45.5	1.2	n.a.	n.a.	
Sample Size—Range^a	399–408	403–413		64–66	62–64	

Source: Teacher and principal surveys.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the 0.05 level, two-tailed test.

n.a.= questions were not asked in the survey.

Table E.15. Educational Background and Certification of Teachers (Percentages)

	Treatment	Control	Impact
Master's Degree or Higher	52.6	57.0	-4.4
Bachelor's Degree From a highly selective or selective college or university ^a	23.1	21.3	1.8
Major			
Elementary education	59.6	52.0	7.6*
Secondary education	3.9	3.5	0.3
Other education	2.0	2.5	-0.4
Subject matter-specific	34.5	42.0	-7.5*
Certification Status			
Regular or standard certificate	89.7	84.7	4.9*
Certified for Current Teaching Position ^b	75.7	76.9	-1.2
Certification Route			
Traditionally certified	88.9	88.4	0.5
National Board Certified	3.8	4.5	-0.7
Number of Teachers—Range^c	343–411	357–415	

Source: Teacher survey.

^aSelectivity of undergraduate institution is defined using Barron's Rankings (2003). Institutions that received the top-three rankings are considered highly selective or selective.

^bTeachers are considered certified to teach their current position if they are certified to teach the grade level, and, for 7th grade teachers, the grade and subject area for the class they are teaching.

^cSample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the 0.05 level, two-tailed test.

Table E.16. Work Experience of Teachers and Principals (Averages Unless Otherwise Noted)

	Treatment	Control	Impact
Teachers			
Teaching Experience			
First year working in the current school (percentage)	16.1	24.5	-8.3*
Years of teaching experience at school	6.6	5.6	1.0*
Years of teaching experience in district	9.0	8.6	0.4
Years of teaching experience	11.9	11.1	0.8*
Nonteaching work experience			
Had a nonteaching job after college (percentage)	24.7	34.1	-9.4*
Years in nonteaching job	1.3	1.9	-0.5*
Number of Teachers—Range^a	402–409	411–415	
Principals			
Years in Current Position	1.7	1.8	0.0
Years in Any Administrative Position at this School	6.2	5.8	0.4
Years in Any Administrative Position	10.7	11.5	-0.8
Number of Principals	66	64	

Source: Teacher and principal surveys.

^aSample sizes are presented as a range based on the data available for each row in the table.

*Impact is statistically significant at the 0.05 level, two-tailed test.

Table E.17. Educational Background and Certification of Teachers, by Teachers' Grade Level (Percentages)

Sample	Treatment-Control Difference		
	All Teachers	Elementary Teachers	Middle School Teachers
Major			
Elementary education	7.6*	10.6*	4.8
Secondary education	0.3	1.1*	-1.0
Other education	-0.4	-1.0	-0.3
Subject matter-specific	-7.5*	-10.7*	-3.5
Certification Status			
Regular or standard certificate	4.9*	3.1	12.1*
Certified for Current Teaching Position ^a	-1.2	-2.1*	2.9
Certification Route			
Traditionally certified	0.5	1.4	-0.9
National Board Certified	-0.7	-1.0	-2.2*
Number of Teachers—Range^b	762–824	502–539	260–285

Source: Teacher survey.

^aTeachers are considered certified to teach their current position if they are certified to teach the grade level, and, for 7th grade teachers, the grade and subject area for the class they are teaching.

^bSample 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

THIS PAGE IS INTENTIONALLY BLANK

