Contract Number:
ED-04-CO-0112 (0012)

Mathematica Reference Number:
06715-400

Submitted to:
Institute of Education Sciences
IES/NCEE
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208
Project Officer: Elizabeth Warner

Submitted by:
Mathematica Policy Research
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Jill Constantine

An Impact Evaluation of the Teacher Incentive Fund (TIF)

Part A

July 13, 2011

CONTENTS

TABLES

SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This package requests clearance from the Office of Management and Budget (OMB) for data collection activities to support a rigorous evaluation of the Teacher Incentive Fund (TIF). This evaluation will include TIF grantees who were awarded funds from the American Recovery and Reinvestment Act (ARRA) of 2009 and the U.S. Department of Education's (ED) fiscal year (FY) 2010 appropriation. The Institute of Education Sciences (IES), within ED, has contracted with Mathematica Policy Research and its partners Chesapeake Research Associates and faculty and staff at the Peabody College of Education at Vanderbilt University to conduct the evaluation.

The main objective of the evaluation is to estimate the impact of differentiated performance-based incentive pay (DPBIP)[1] on student achievement and the mobility and retention of teachers and principals. The evaluation design is an experiment in which researchers will randomly assign schools within a district to either a treatment or control group. The treatment schools will implement educator DPBIP as part of a performance-based compensation system (PBCS). Control schools will implement the same non-differentiated components of the PBCS program and a one percent across-the-board bonus, but will not implement any type of DPBIP for the duration of the TIF grant. We will compare student achievement and other outcomes between the treatment and control schools to estimate the impact of DPBIP compared to the one percent bonus.

The Notice of Final Priorities (NFP) for the TIF grants, published in the *Federal Register* on May 21, 2010, announced two competitions for grants to be awarded in 2010—the TIF main competition and the TIF evaluation competition; applicants applied to one or the other competition. Successful applicants for the evaluation competition received an "evaluation grant" that includes an additional financial award to fund TIF program activities, including some activities that are not eligible for funding under the main competition. Grantees awarded an evaluation grant had to demonstrate their ability and willingness to meet the grant requirements, which included the main competition requirements plus additional ones specific to the evaluation. In particular, evaluation grantees agreed to cooperate with data collection activities required for the national evaluation, identified the schools that will participate in the national evaluation, and agreed to allow those schools to be randomly assigned to either the treatment or control group. Both main and evaluation grants are for five years.

This is the second submission of a two-stage clearance request for the evaluation. The first package (approved October 18, 2010, under OMB #4285) requested clearance to ensure that grantees' program designs and implementation are consistent with the requirements for a rigorous evaluation of the TIF, and if necessary, recruit grantees for the evaluation. This second package requests clearance to collect data that will support the full-scale study.

We believe it is important to note that our eventual data collection plans will differ in two ways from those for a study of TIF grantees being conducted by the Policy and Program Studies Services (PPSS) in the Office of Planning, Evaluation and Policy Development at ED. First, the two data

[1] For this document, DPBIP refers to the differentiated incentive pay portion of a grantee's PBCS. DPBIP programs provide bonuses for highly effective teachers and principals, where effectiveness is based on student achievement growth, observations, and any other criteria included in the district's PBCS.

collection efforts target different respondents. The PPSS study includes grantees from the FY2007 awards while participants in the current study received their grants in FY2010, and the two studies target *different* schools and/or educators. Second, the focus and design of each study is different. The PPSS evaluation is an implementation study. This evaluation uses a rigorous experimental design in which schools are randomly assigned to either a control or treatment group to estimate the impact of DPBIP on student achievement and educator mobility and recruitment.

Part A. Justification

1. Circumstances Necessitating the Collection of Information

a. Statement of Need for a Rigorous Evaluation of TIF

The specific legislation necessitating and funding this data collection is the ARRA, Division A, Title VIII, Pub. L. 111–5 and Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Act, 2010, Division D, Title III, Pub. L. 111–117. The ARRA requires that ED, to the extent possible, conduct a rigorous national evaluation to assess the impact of PBCS, supported by ARRA funds, on student achievement and educator recruitment and retention in high-need schools and hard-to-staff subjects. This evaluation would meet this requirement.

Local educational agencies (LEAs) use TIF grants to implement performance-based teacher and principal compensation systems in high-need schools. ARRA requires that the funding be used to promote effective school reform in several priority areas. These priorities include increasing teacher effectiveness, achieving equity in the distribution of high-quality teachers, and turning around the lowest performing schools. TIF requirements address these priorities.

Teacher quality is a critical input to student learning, but little is known about how to develop a strong teacher workforce (Rivkin et al. 2005; Rockoff 2004). Researchers have examined strategies to identify, attract, retain, and develop good teachers, including alternative preparation (Decker et al. 2004; Constantine et al. 2009), certification (Tuttle et al. 2009), and in-service training and professional development (Glazerman et al. 2006, Garet et al. 2008, Yoon et al. 2007). However, little is known about incentive compensation programs that tie teacher pay to student performance. Do these programs boost student achievement by attracting and retaining effective teachers and motivating all teachers to improve performance? Which types—for example, school- or individual-based programs or mixed programs (a combination of the two)—are most effective? And what challenges do districts face in implementing these programs?

To assess the overall effectiveness of TIF projects and the effectiveness of particular program models and features, ED has contracted for an evaluation of DPBIP that will be implemented by the most recent round of grant recipients. This evaluation will provide important evidence on how changes to the traditional compensation systems for educators may be able to (1) improve student performance in high-need schools and/or (2) bring about desirable changes, such as the presence of more highly effective educators in high-need schools. Results of this evaluation will provide educators, policymakers, and researchers with critical information on educator compensation reform, the effect of performance-based educator compensation on student achievement, and other aspects of PBCSs associated with student achievement.

b. Research Questions

The study's research questions are:

- What is the impact of DPBIP on student achievement and educator mobility and recruitment?
- Is a particular type of DPBIP model—for example, school- or individual-based or mixed models—associated with greater growth in student achievement?
- Are other key program features correlated with student and educator outcomes?
- What are the experiences and challenges of districts when implementing these programs?

c. Study Design

To answer the first research question, this study will use an experimental design—study schools within a district will be randomly assigned to either a treatment or control group. Random assignment is considered the "gold standard" for social policy evaluations. More than any other approach, it minimizes the chance that any observed differences in outcomes between the study groups are due to unmeasured, pre-existing differences between members of these groups. In the random assignment design, the simple difference between outcomes in treatment and control schools within each district is an unbiased estimate of the impact of the district's DPBIP component.

Both treatment and control schools will implement the same non-DPBIP components of their program. However, only treatment schools will include a DPBIP component, while control schools will provide an across-the-board one percent educator bonus. Control schools will not be permitted to implement a DPBIP component for the duration of the TIF grant.

Treatment schools must implement both teacher and principal DPBIP components that measure effectiveness using gains in student academic achievement and classroom evaluations conducted multiple times during each school year. Teacher incentive models may be individual-based, group-based, or mixed models.

Since we will not randomly assign schools to specific program features (program features differ among grantees), the study will use nonexperimental analyses to address the other research questions. To the extent possible, the study will examine the correlation between different types of DPBIP models and student and educator outcomes. The ability to separately analyze different DPBIP models will depend on the number and type of model(s) implemented by the grantees. The study will also examine the association of other key program features, such as how heavily the DPBIP model weights growth in student achievement with student achievement and educator outcomes.

The ability of the study to detect differences between the treatment and control groups depends, in large part, on the sample sizes. The study will include approximately 250 schools and students. It is designed to detect student achievement gains of 0.09 of a standard deviation. Although this may be a larger effect than can be obtained in the first year or two of the program, if DPBIP is effective in retaining and attracting effective teachers as well as improving performance among all teachers, improvement in student achievement should increase over time as educators

observe bonuses received by colleagues. In addition, relatively small gains could be realized each year, contributing to larger effects after three or four years of implementation.

As part of the evaluation, and to address the research questions, Mathematica will:

- Collect principal and teacher contact information for the study team to contact respondents who may change schools during the course of the study.
- Collect student records data to estimate the impact of DPBIP on student achievement.
- Collect administrative data on principals and teachers to track their mobility and recruitment.
- Use principal and teacher surveys to describe their understanding of and experiences with DPBIP, supplement district mobility data, and obtain background information.
- Use district surveys and interviews to describe experiences and challenges of districts when implementing the incentive programs.

d. Data Collection Needs

This study includes several data collection efforts, described below and summarized in Table 1 below. Data will be collected from the districts and schools participating in the evaluation.

Teacher and principal contact information. At the beginning of the 2011–2012 school year, we will administer a contact form to all principals in the study schools and a subsample of teachers who will be requested to complete the teacher survey if we cannot obtain this information from administrative records (Appendix A). For these schools/districts, we will administer the same form in fall 2012, 2013 and 2014, to new principals who transfer into study schools and new teachers who fill positions of those who previously completed the teacher survey. This form will request detailed contact information such as the respondent's telephone number and permanent address. It will also include the study survey director's telephone number and email address for teachers or principals to contact, if necessary. We will mail the forms in one package to the school principal and ask him or her to distribute to the teachers. The principal will also collect the completed forms and mail them to Mathematica in a postage-paid envelope.

District survey. We will administer a survey in three rounds to all 2010 TIF main and evaluation districts (Appendix B). The first round, to be administered in fall 2011, will request information on specific features of the incentive program, if changes were made to the program since grantees submitted their application, approaches districts used to obtain buy-in as well as any compromises they had to make, and expectations for educator incentive payouts. The second round, to be administered in fall 2012, will ask about any changes districts made in their system, reasons for the changes, and experiences and outcomes from the first year of program implementation. The third round will be administered in fall 2014 and will focus on districts' experiences over the longer period and their plans for sustaining the incentive policies. The survey seeks to contrast how the districts' programs were planned, implemented, and sustained. This package includes the instrument that will be used in the first round of data collection. Later rounds will be very similar in format and structure. However, we expect that responses from the first round (and later the second round) may inform revisions to subsequent rounds.

We will mail the 30-minute hard copy questionnaire to each district representative. The mailing will contain a cover letter and district questionnaire. The letter, which will be on ED stationary and signed by an ED official, will describe the study and its objectives and the need for districts' participation, address issues of confidentiality, and provide a senior study member's contact information for questions or concerns. Districts will be asked to complete a hard copy questionnaire and mail it to Mathematica in a postage-paid envelope.

Table 1. Data Collection Needs

Instrument	Data Need	Respondent	Mode	Schedule
Principal and teacher contact form	Personal contact information to enable contact if educator leaves school during study	Teachers and principals	Hard copy or electronic if available	Fall 2011 Fall 2012, 2013, 2014 (new teachers)
District questionnaire	Specific program features, changes made to program, and how district obtained buy-in	District staff	Hard copy, phone follow-up	Fall 2011, Fall 2012, Fall 2014
Principal questionnaire	Background characteristics, mobility, and knowledge and perceptions of incentives	Principals	Web with email, hard copy and phone follow-up	Spring 2012, 2013, 2014, 2015
Teacher questionnaire	Background characteristics, mobility, and knowledge and perceptions of incentives	Teachers	Web with email, hard copy and phone follow-up	Spring 2012, 2013, 2014, 2015
Principal and teacher administrative data letter	Educator retention, school assignment, background characteristics, standardized test scores	District staff	Electronic or hard copy	Summer/fall 2011, 2012, 2013, 2014, 2015
Student administrative records letter	Reading and math standardized test score data for current and prior school year	District staff	Electronic or hard copy	Summer/fall 2012, 2013, 2014, 2015
	Demographic and socioeconomic characteristics			
District interview protocol	Detailed information on program, implementation experiences, and other school improvement efforts	District staff	Telephone semi-structured interviews	Spring/summer 2012, 2013, and 2015

Principal survey. A 30-minute web-based survey will be administered to all principals in four waves—spring 2012, 2013, 2014, and 2015 (Appendix C). We will administer later surveys to the same principals even if they have left the school, as well as new principals in study schools. The principal survey will ask about their background characteristics, mobility, the school's hiring practices, and knowledge and perceptions of incentives.

Teacher survey. Administered to a sample of teachers, the teacher survey (Appendix D) will be similar to the principal survey regarding mode of administration and follow-up and length of questionnaire. As with principals, in follow-up years, we will administer surveys to the same teachers even if they have left the school, as well as new teachers in study schools. The survey will collect information on teachers' educational and professional background, professional development experiences, teaching and leadership responsibilities, satisfaction with various aspects of their schools, salary and other sources of compensation, and understanding of their school's PBCS.

For both principal and teacher surveys, we will first contact the sample members by email or cover letter (if email is not available or invalid). The initial correspondence will include a description of the study and survey, a link to the website address and instructions on accessing the survey, and a

unique username and password. The email will explain the importance of participation, address confidentiality, and provide a toll-free telephone number and email address for questions or concerns. Nonrespondents, whom we will contact by email, telephone, or a remailing, will have the additional option of providing answers either over the telephone or by completing a hard copy version of the questionnaire.

Principal and teacher administrative data. In fall 2011, and annually through 2015, we will collect data from districts on the hiring, movement between schools, and attrition of principals and teachers participating in the study. We will also attempt to obtain information about the start and end dates of school assignments for these staff, as well as any available background characteristics such as age, sex, race/ethnicity, certifications, degrees, years of teaching experience, and scores on licensure or certification tests. In addition, we will collect several indicators of teacher and principal effectiveness and data on the actual payouts received by staff in recognition of their accomplishments. We will collect these data by the following means:

- Annual listings of principals and teachers (with personnel ID code, school, and grade if applicable) who are eligible for performance pay and the maximum amounts for which they are eligible.
- Annual listings of principals and teachers (with personnel ID code, school, and grade if applicable) who actually receive performance pay and the amounts that they receive.
- Annual data on performance measures received by principals and teachers in treatment and control schools (with personnel ID code, school, and grade if applicable). To the extent possible, performance measures should be separated into those based on observations of classroom or school practices, student achievement and growth, and other performance criteria.

Although we prefer to receive the data in an electronic format, we will use data in whatever form is most convenient for each district. We will send letters to the districts, specifying the specific data elements requested (Appendix E).

Student records data. We will request standardized math and reading test scores for all students in study schools in spring 2012, 2013, 2014, and 2015. We will also request scores from the year prior to the current study year if those scores have not been previously obtained. In addition to test scores, we will request that the district data on student characteristics such as sex, race/ethnicity, date of birth, grade, whether they are repeating a grade, eligibility for free- or reduced-price lunch, English language learner status, and mobility within the district. Where possible, we will also request student achievement scores in math and reading, linked to the appropriate teacher. We will send the district a letter specifying the data requested (Appendix F).

District interviews. In spring 2012, 2013, and 2015, we will conduct semi-structured telephone interviews with a district official who is familiar with the TIF evaluation grant program. The interview protocol is designed to collect detailed information on each district in a format that will allow for standardized follow-up questions depending on the response given to a specific item. The interview will address topics such as program implementation experiences and other ongoing school improvement efforts. The protocol for the initial administration in 2012 is included in Appendix G; subsequent administrations will be tailored to address issues pertinent to the administration of the grants following start-up years.

e. Study Activities and Data Collection Timeline

This clearance request pertains to the collection of principal and teacher contact information (Appendix A); administration of the district survey (Appendix B), principal survey (Appendix C), and teacher survey (Appendix D); collection of the district administrative records on principals, teachers, and students in the study (Appendices E and F); and administration of a district interview (Appendix G). The evaluation will be completed in seven years. Table 2 shows the schedule of data collection activities and the overall evaluation timeline.

Table 2. Schedule of Major Study Activities

Activity	Fall 2010	Fall 2011	Spring 2012	Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014	Spring 2015	Fall 2015	Spring 2016
Solidify grantee participation	X										
Provide technical assistance to grantees	X	X	X	X	X	X	X	X	X		
Collect principal and teacher contact information		X		X		X		X			
Conduct district survey		X		X				X			
Conduct principal survey			X		X		X		X		
Conduct teacher survey			X		X		X		X		
Collect principal and teacher records data from districts		X		X		X		X		X	
Collect student records data from districts				X		X		X		X	
Conduct district interviews			X		X				X		
Prepare first report					X						
Prepare second report							X				
Prepare third report									X		
Prepare fourth report											X

2. Purposes and Uses of Data

Data for the evaluation of TIF programs will be collected and analyzed by Mathematica and its partners, Chesapeake Research Associates and the Peabody College of Education at Vanderbilt University. This work will be conducted under contract number ED-04-CO-0112. The data to be collected will be obtained from participants' contact information, district administrative records, TIF district interviews, and surveys of teachers, principals and districts. The data will be used to address the research questions as shown in Table 3.

Table 3. Research Questions and Data Collection Methods

Research Question	Data Sources
1. What is the impact of DPBIP on student achievement and educator mobility and recruitment?	• District administrative records • Principal survey • Teacher survey
2. Is a particular type of DPBIP model—for example, school- or individual-based or mixed programs—associated with greater growth in student achievement?	• District administrative records • Principal survey • Teacher survey • District survey • District interviews
3. Are other key program features correlated with student and educator outcomes?	• District administrative records • Principal survey • Teacher survey • District survey • District interviews
4. What are the experiences and challenges of districts when implementing these programs?	• District survey • Principal survey • Teacher survey • District interviews

- **Principal and teacher contact forms.** The information collected via this form will be used to contact participants who leave the school during the grant period so we can ask them to complete their respective surveys.

- **District survey.** We will use the data from three district surveys to examine the association between impacts and key program features. Data from the first survey will be used to examine specific features of the incentive program and to understand approaches districts used to obtain buy-in and compromises they had to make. We will use information from the second survey to explore districts' experiences in the first year of program implementation and changes they had to make. Finally, data from the third survey will be used to describe districts' experiences since implementing the TIF program and ascertain their plans for sustaining the program. Data from the district surveys will be used to answer research questions 2, 3, and 4.

- **Principal survey.** The principal survey will be used to assess hiring practices, classroom assignments, knowledge and perceptions of the TIF program in the study schools, how this may change over time, and to supplement administrative data to be obtained from district records. The principal survey can also provide important insight on their motivation for remaining, leaving, or entering a study school. Data from the principal survey will be used to answer research questions 1, 2, 3, and 4.

- **Teacher survey.** The teacher survey will be used to assess knowledge and perceptions of the PBCS in the study schools and how this may change over time, and to supplement administrative data to be obtained from district records. The teacher survey can also provide important insight on teachers' motivation for remaining, leaving, or entering a

study school. Data from the teacher survey will be used to answer research questions 1, 2, 3, and 4.

- **Principal and teacher administrative data.** These data will be used to estimate the impacts of DPBIP on educator mobility and recruitment. The data will also allow us to examine the association between educator characteristics and student and educator outcomes, and to describe the educator sample. These data will be used to answer research questions 1, 2 and 3.
- **Student records data.** We will use existing state or district test score data to estimate the impact of DPBIP on student achievement, the key outcome of interest. Information on students' demographic and socioeconomic characteristics and their achievement test scores prior to the study school year will be used to describe the students in the study and to develop more precise impact estimates. To the extent possible, we will use student-teacher linked data to estimate teachers' value-added score to better understand mobility of high- and low-performing educators. Data obtained from student records will be used to address research questions 1, 2 and 3.
- **District interview.** The semi-structured district interviews will allow us to collect more in-depth information than that collected from the survey, and to follow up for clarification if necessary. We will use this detailed information to more thoroughly understand each program's context, implementation strategy, and challenges. Data from the district interviews will be used to answer research questions 2, 3, and 4.

The overall purpose of this evaluation is to estimate the impacts of DPBIP on student achievement and educator mobility and recruitment in high-need schools. The findings from this study will provide important evidence for school districts and policymakers on the impacts of DPBIP on students, teachers, and school principals. If possible, this evaluation may provide policymakers and school districts with valuable information on the relative effectiveness of individual-based versus group-based compensation systems. The study will also provide important insight into the impacts of other key program aspects of DPBIP models, as well as how districts may overcome common implementation challenges. Study findings will be presented in four annual reports, beginning fall 2013. In addition, the data collected by the evaluation will be available as restricted-use data files that will serve as a valuable resource for other researchers.

3. Use of Technology to Reduce Burden

The data collection plan is designed to obtain reliable information in an efficient way that minimizes respondent burden. We will set up a toll-free telephone number and email address specific to the study so that participants with questions can easily contact the research team. As much information as possible will be gathered from existing data sources, such as TIF grant application packets submitted by awardees and electronic files provided by districts. If it is too burdensome or not possible for a district to provide data in electronic format, we will provide clear instructions on how to submit copies of the relevant information in hard copy form, to be coded by the study team. Some data, however, can only be obtained directly from principals, teachers, and districts.

A web-based survey will be the primary mode of data collection for teachers and principals in the study. Respondents will also have the option of completing a self-administered hard copy questionnaire or providing answers to a trained interviewer over the telephone. The web-based

survey will enable respondents to complete the survey at a location and time of their choice, and its automatic editing system will reduce the number of response errors.

For participants who do not return contact forms, or those whose email addresses are invalid, we will search school or district websites to obtain email addresses. Using email to follow up with nonrespondents will also offer an additional convenient option for respondents. Email reminders will include a link to the survey website and a username-password combination, as well as an attached PDF of the survey if respondents choose to complete a hard copy version.

A district representative familiar with the TIF program will complete questionnaires in hard copy form. For nonresponse follow-up, we will also offer respondents the opportunity to complete the survey over the telephone with a trained telephone interviewer. The study team considered other modes of administering the district survey, such as computer-assisted telephone interview (CATI) or a web-based survey. However, because of the relatively small sample size, the predicted cost of developing these methods outweighed the expected benefits.

We will conduct the district interviews by telephone. This mode of data collection is appropriate for the conversational exchange necessary to obtain answers to the open-ended questions, and to allow probing for more detail than a self-administered survey can provide.

4. Efforts to Avoid Duplication of Effort

The data collection plan avoids unnecessary collection of information from multiple sources. For example, the study will obtain preliminary information about grantees from existing district databases, grant applications, and administrative records. The preliminary information is helpful in examining factors such as the variation of program features, including the size and distribution of award, how performance awards compare in size to other incentives, and the relative weighting of school- or individual-based criteria. These factors will help guide the subgroup and correlational analyses.

Although the Policy and Program Studies Services (PPSS) in the Office of Planning, Evaluation and Policy Development at ED is conducting a study of TIF grantees, there are important differences between the two evaluations. First, the two data collection efforts target different respondents. The PPSS study includes grantees from the FY2007 awards whereas participants in the current study received their grants in FY2010. While some grantees have both FY2007 and FY2010 awards, each award covers *different* schools and/or educators, thus there is no overlap at the school level. Furthermore, we will coordinate with PPSS to avoid requesting duplicate information from participants.

Second, the focus and design of the two studies are different. The PPSS evaluation is an implementation study which aims to describe districts' program features and implementation experiences.Our evaluation uses a rigorous experimental design in which schools are randomly assigned to either a control or a treatment group to estimate the impact of DPBIP on student achievement and educator mobility and recruitment.

5. Methods to Minimize Burden on Small Entities

The primary entities for the study are TIF school districts, schools, principals, and teachers. We will minimize burden for all respondents by requesting only the minimum data required to meet

study objectives. Burden on respondents will be further minimized through the careful specification of information needs. We will also keep our data collection instruments short and focused on the data of most interest, and we will speak with relatively few respondents in person. Sample sizes and data requirements for each respondent group were determined by careful consideration of the information needed to meet the study objectives, and were reviewed by the study's technical working group (TWG).

6. Consequences of Not Collecting Data

The data collection plan described in this submission is necessary for ED to conduct a rigorous national evaluation of the TIF and to understand the effectiveness of this education reform strategy. Collecting these data will allow us to examine the range of performance-based compensation systems and to answer pressing policy questions about how DPBIP affects student achievement and how grant recipients design, communicate, and implement TIF programs.

The consequences of not collecting specific data are outlined below.

- Without the information from the **principal and teacher contact forms**, the study will lose track of sample members when they change schools or leave the profession. This is especially critical if an educator leaves a study school or district.
- Each wave of the **district survey** targets different aspects of the program: specific features of districts' PBCS, if and how these features changed over time, how districts obtained buy-in, their experiences, and plans to continue their incentive policies. Without administering the district survey, and in multiple waves, we will not be able to capture these key program features and their impact on student achievement and educator mobility.
- Without the **principal and teacher surveys,** we will not know if educators understood the incentive policies, if their choice to stay in, move to, or move from a school was motivated by the incentives. We will also be unable to examine schools' hiring practices and classroom assignments, two factors that may be influenced by the TIF program. Impacts in the second and subsequent years of the implementation of the DPBIP may be larger than those in the first year. Administering the surveys in multiple waves will allow us to examine educators' experiences and perceptions of the programs over time.
- Without **principal and teacher records data**, it will be more difficult to verify educators' school assignment and track their mobility. Furthermore, without this data we will not be able to compare characteristics between principals and teachers in the treatment and control schools, or to examine whether staff characteristics are associated with student achievement growth or eductors' mobility decisions.
- Without **student records data**, we will have to administer assessments to students in place of using their district or state math and reading test scores to measure student achievement. Without the data on student characteristics, we will not be able to fully describe the study sample and verify the effectiveness of the random assignment.
- Without the **district interviews**, we will not be able to follow up on information obtained from the surveys to obtain a more thorough understanding of the districts' programs and experiences, or to fully understand any other related school reform

initiatives within the district that may affect the impact of DBPIP in the study schools. Multiple waves are necessary as a detailed follow-up to each district survey.

7. Special Circumstances

There are no special circumstances associated with this data collection.

8. *Federal Register* Announcement and Consultation

a. *Federal Register* Announcement

The 60-day notice to solicit public comments was published in Volume 75, Number 77, page 22387 of the Federal Register on April 21, 2011. No public comments were received. . The 30-day notice will be published to solicit additional public comments.

b. Consultation Outside the Agency

In formulating the evaluation design, the study team sought input from the technical working group (TWG), which includes some of the nation's experts in teacher compensation, evaluation methodology, and education policy. We will continue to consult with the TWG throughout the study on other issues that would benefit from their input. Table 4 lists the TWG members.

Table 4. Technical Working Group Members

Name	Title and Affiliation	Expertise
Anthony Milanowski	Assistant Scientist, University of Wisconsin	Teacher compensation
Richard Murnane	Professor of Education, Harvard Graduate School of Education	Teacher compensation and teacher quality
Jacob Vigdor	Professor of Public Policy and Economics, Duke University	Teacher compensation, teacher quality, and evaluation methodology
Dan McCaffrey	Senior Statistician, RAND Corporation	Value added and evaluation methodology
Robert Meyer	Research Professor, University of Wisconsin	Value added
Jeffrey Smith	Professor of Economics, University of Michigan	Teacher quality/methodology
James Kemple	Director of Research Alliance for NY City Schools, Research Professor, New York University	Teacher quality/methodology
David Heistad	Executive Director of Research, Evaluation and Assessment, Minneapolis Public Schools	Program evaluation, value-added in teacher compensation systems
Carla Stevens	Assistant Superintendent, Research and Accountability, Houston Independent School District	Accountability, student assessment, program evaluation, and performance pay models

c. Unresolved Issues

There are no unresolved issues.

9. Payments or Gifts

Incentives for principals and teachers. Incentives have been proposed for the principal and teacher surveys to partially offset respondents' time and effort in completing the surveys. We

propose offering a $20 incentive to an educator each time he or she completes a questionnaire so as to acknowledge the 30 minutes required to complete each questionnaire. This proposed amount is within the incentive guidelines outlined in the March 22, 2005 memo, "Guidelines for Incentives for NCEE Evaluation Studies," prepared for OMB.

Incentives are also proposed because high response rates are needed to make the survey findings reliable, and we are aware that teachers and principals are the targets of numerous requests to complete surveys on a wide variety of topics from state and district offices, independent researchers, and the Department of Education. Although some districts will have solicited buy-in from teachers to participate in the evaluation, our recent experience with numerous teacher surveys supports our view that obtaining teacher buy-in does not guarantee teachers will devote the time it takes to complete a survey, and monetary incentives increase the likelihood of cooperation of school staff.

The study will not give incentives to districts for completing an interview or a survey, or for providing administrative records data.

10. Assurances of Confidentiality

Mathematica and its research partners will conduct all data collection activities for this study in accordance with relevant regulations and requirements, which are:

- The Privacy Act of 1974, P.L. 93-579 (5 U.S.C. 552a).
- The Family Educational and Rights and Privacy Act (FERPA) (20 U.S.C. 1232g; 34 CFR Part 99).
- The Protection of Pupil Rights Amendment (PPRA) (20 U.S.C. 1232h; 34 CFR Part 98).
- The Education Sciences Institute Reform Act of 2002, Title I, Part E, Section 183

The research team will protect the confidentiality of all data collected for the study and will use it for research purposes only. The Mathematica project director will ensure that all individually identifiable information about respondents will remain confidential. All data will be kept in secured locations and identifiers will be destroyed as soon as they are no longer required. All members of the study team having access to the data will be trained and certified on the importance of confidentiality and data security. When reporting the results, data will be presented only in aggregate form, such that individuals and institutions will not be identified. Included in all voluntary requests for data will be the following statement:

> Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law. Additionally, no one at your school or in your district will see your responses. While your participation in this study is voluntary, it is very important that you complete the questionnaire.

For those instruments where data collection is required as a condition of their evaluation grant, all grant required requests for data will include the following statement:

> Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law. Additionally, no one at your school or in your district will see your responses. Participation or cooperation with this activity is a condition of your grant (EDGAR: part 75.591, Authority: 20 U.S.C. 1221e–3 and 3474).

The following safeguards are routinely employed by Mathematica to carry out confidentiality assurances, and they will be consistently applied to this study:

- All Mathematica employees sign a confidentiality pledge (Appendix H) that emphasizes the importance of confidentiality and describes employees' obligations to maintain it.
- Personally identifiable information (PII) is maintained on separate forms and files, which are linked only by sample identification numbers.
- Access to hard copy documents is strictly limited. Documents are stored in locked files and cabinets. Discarded materials are shredded.
- Access to computer data files is protected by secure usernames and passwords, which are only available to specific users.
- Sensitive data is encrypted and stored on removable storage devices that are kept physically secure when not in use.

Mathematica's standard for maintaining confidentiality includes personnel training regarding the meaning of confidentiality, particularly as it relates to handling requests for information, and providing assurance to respondents about the protection of their responses. It also includes built-in safeguards concerning status monitoring and receipt control systems.

11. Justification for Sensitive Questions

Some respondents may consider their contact information to be sensitive. This information is necessary in order to limit possible sample attrition that could result from respondents changing schools or professions.

The principal and teacher surveys will ask for demographic information (ethnicity, race, year of birth) and information about respondents' educational and professional background. Data on these topics are important to help us understand if there is an association between student achievement, educator outcomes, and educator characteristics. Questions used to obtain personal background information have been asked frequently in other surveys and were pretested for this study, with the pretest sample of teachers and principals reporting no concerns.

To address concerns about disclosing personal information, all cover letters and questionnaires will clearly state that all responses will be treated as confidential, that participation is voluntary, and that failure to provide some or all requested information will not affect the respondent's professional status in any way. The questions will also be worded in a sensitive, nonjudgmental manner.

Some demographic information about the students (for example, qualification for free- or reduced-price lunch or special education status) or their test scores may be sensitive. Demographic

information is important to control for any differences in the characteristics of students in the classes that may have arisen by chance. Test score data is essential for this evaluation because student achievement is the primary outcome of interest. These scores will be linked to the data file by each respondent's unique, study-generated identification number. After this linking process, personal identifiers, such as a student's name, school identification number, and date of birth, will be removed.

There are no questions of a sensitive nature in the district survey or interview.

12. Estimates of Hours Burden

Table 5 provides an estimate of time burden for the data collections, broken down by instrument and respondent. These estimates are based on our experience collecting administrative data from districts, administering surveys to school principals and teachers, and conducting telephone interviews with district representatives.

Table 5. Estimated Response Time for Data Collection

Respondent/ Data Request	Number of Targeted Respondents	Expected Response Rate (%)	Number of Respondents	Unit Response Time (Hours)	Total Response Time (Hours/Year)	Total Burden Time (Hours)
Districts[a]						
Student records data (4 times)	15	100	15	8.0	120	480
Principal and teacher records data (4 times)	15	100	15	8.0	120	480
Principals						
Principal contact information[b] (once)	362	90	326	0.08	26	26
Principal surveys (4 times)	250	90	225	0.5	113	450
Teachers						
Teacher contact information[c] (once)	3,500	85	2,975	0.08	238	238
Teacher surveys (4 times)	2,000	85	1,700	0.5	850	3,400
Districts						
Surveys (3 times)	186	80	149	0.5	74.5	224
Interviews (3 times)	15	100	15	0.75	11.25	34
Total						5,332

Annual number of respondents and responses for the 3 years of this collection are 3220 and the total annual burden hours for this collection are 1377 burden hours.

[a]Depending on the grantee, administrative records data may be provided by another source, for instance the state or grantee.

[b]We assume 15 percent of the principals will be replaced each year.

[c]We assume 25 percent of the teacher sample will be replaced each year.

The total of 5,332 hours covers all four years of the evaluation, and includes the following efforts: up to 16 hours, annually for four years, for each of the 15 districts to collect and assemble administrative records on students, principals, and teachers participating in the evaluation; 30 minutes, annually for four years, for 225 principals (90 percent of the anticipated 250 principals in the sample) to complete the principal survey; 30 minutes, annually for four years, for 1,700 teachers (85 percent of the anticipated sample of 2,000 teachers in the sample) to complete the teacher survey; 30 minutes, annually for three years, for 149 district representatives (80 percent of the 186 districts participating in the study) to complete a district survey; and 45 minutes for the 15 districts participating in the evaluation to complete a telephone interview each year for three years. Annual number of respondents and responses for the 3 years of this collection are 3220 and the total annual burden hours for this collection are 1377 burden hours.

13. Estimates of Cost Burden to Respondents

There are no direct costs for respondents.

14. Annualized Costs to the Federal Government

The estimated annual cost of the study to the federal government is $1,714,286. The total cost of the seven-year study is $12 million, which includes recruiting grantees, districts, and schools; designing and administering data collection instruments; processing and analyzing data; and preparing reports.

15. Reasons for Program Changes or Adjustments

There is an overall program change increase of 1130 burden hours. This program change is a result of the burden hours from this second phase (1377) being added and the burden hours (247) of the first phase (the recruitment phase) being eliminated since they will be completed by the time this second phase is approved.

16. Plans for Tabulation and Publication of Results

a. Tabulation Plans

Our tabulation plans include four sets of analyses aligned to the research questions. Random assignment of schools within a district to a treatment group that will implement DPBIP or to a control group not allowed to do so for the duration of the TIF grant is an ideal design for assessing overall effectiveness. Our primary impact analysis will exploit this experimental design to provide rigorous estimates of the impact of DPBIP on student achievement and teacher/principal mobility and recruitment. Additional nonexperimental analyses are designed to estimate the relative effectiveness of individual-based versus group-based or mixed incentive programs, explore the association of other key program features with student achievement and teacher/principal outcomes, and to learn about districts' implementation experiences and challenges.

Estimating the overall impact of DPBIP. With this experimental design, the simple differences between mean outcomes in the treatment and control schools should yield unbiased estimates of the impacts of DPBIP. However, the precision of the estimates can be improved by using regression procedures to control for student, teacher, or school baseline characteristics that

may explain some of the variation in outcomes not related to the treatment itself. These characteristics may include student controls, such as test scores from the year before TIF implementation; gender, race/ethnicity, free- or reduced-price lunch eligibility, special education status, and English learner status; teacher controls, such as demographic characteristics, age, experience, and educational background; and school-level averages of the student or teacher characteristics. Regression procedures also enable us to adjust for any differences between treatment and control groups in these baseline characteristics that happen to arise due to chance or sample attrition. The regression model must be flexible enough to include the full range of programs and generate estimates of district-specific impacts, which can then be aggregated to produce an overall estimate. We will therefore estimate variations of the following model for the outcome y_{ijk} of individual (student or teacher) i in school j within district k:

$$(1) \quad y_{ijk} = \mathbf{R_{jk}}\boldsymbol{\alpha} + \sum_{k=1}^{K} \beta_k (T_{jk} \times G_k) + \mathbf{X_{ijk}}\boldsymbol{\delta} + \mathbf{Z_{jk}}\boldsymbol{\gamma} + u_{jk} + \varepsilon_{ijk}$$

where $\mathbf{R_{jk}}$ is a vector of indicators for combinations of grade levels and randomization strata; $\boldsymbol{\alpha}$ is a vector of grade-by-strata fixed effects; T_{jk} is a treatment indicator; G_k is a dummy variable for district k; β_k is the impact of DPBIP in district k; $\mathbf{X_{ijk}}$ is a vector of baseline individual characteristics with coefficient vector $\boldsymbol{\delta}$; $\mathbf{Z_{jk}}$ is a vector of baseline school-level characteristics with coefficient vector $\boldsymbol{\gamma}$; u_{jk} is a random school effect; and ε_{ijk} is a random individual error term. The district-specific impacts of performance pay, β_k, are the key coefficients of interest in equation (1). We will estimate equation (1) with ordinary least squares (OLS) using Huber-White ("sandwich") standard errors that account for school-level clustering.

Our primary interest is in the overall, average impact of DPBIP in the full study sample. To estimate the average impact of DPBIP on schools in the study, we will take a weighted average of the estimated district-specific effects, $\hat{\beta}_k$, with weights equal to the number of treatment and control schools within each district. The standard error of the average impact estimate can be calculated from the estimated variances and covariances among the district-specific impacts from equation (1).

The evaluation includes four years of analyses. Impacts in the second and subsequent years of the implementation of the DPBIP may be larger than those in the first year for several reasons. First, changes in educator effort and the composition of the teaching staff at treatment schools may be more pronounced after educators observe the payments from earlier years. Also, if educators improve their performance over time, in years 2 through 5 of the grant, some students will have had multiple years of exposure to the treatment. For these reasons, equation (1) will be estimated separately for assessing impacts for each year of implementation, as well as cumulative impacts.

The impact of DPBIP on the outcomes of interest—student achievement and educator mobility and recruitment—will be estimated with a variant of equation (1). Student achievement outcomes are math and reading scores from spring 2012, 2013, 2014, and 2015 state or district assessments. Because tests will differ across states, grade levels, and subjects, we will convert raw scale scores to z-scores (raw scores minus the mean score divided by the standard deviation of scores on that test among students in that grade and state) in order to scale the outcome variable

comparably across all students in the sample. Using district records, we will measure teacher retention as a dichotomous outcome for whether or not the teacher returns to work in the grantee site and/or in his or her initial school in fall of 2011 and continue to do so annually through 2015. Because the retention outcome is dichotomous, we will estimate the probit model analog of equation (1). Annual school-level teacher data from study schools in fall, 2011 through fall, 2015 (from district records) and spring 2012, 2013, 2014, and 2015 (from the principal and teacher surveys) will be analyzed as outcomes to examine impacts on the composition of the teaching staff. If available from administrative records, the quality of applicants who apply to teach in study schools for school years 2012–2013, 2013–2014, 2014–2015, and 2015–2016 will also be analyzed, including the total number of applicants, average experience level, percentage of applicants who have teaching experience, and the selectivity of the college from which they graduated. Equation (1) can be aggregated to the school level for the analysis of composition outcomes.

To better understand mobility of high- and low-performing principals and teachers, for grantees where we can obtain or calculate a measure of staff effectiveness, we will also estimate a model of transitions that includes a teacher or school measure of effectiveness, and interactions of this measure with treatment indicators in the set of independent variables. The coefficients on the effectiveness measure by treatment interactions provide an estimate of whether differences in retention between highly effective and less effective principals or teachers are more or less pronounced in treatment versus control schools. Since high- and low-performing teachers are not being randomly assigned to treatment and control schools, and estimates of their effectiveness may be endogenous if DPBIP induces greater teacher effort, these estimates are nonexperimental and will need to be interpreted with caution. Wherever possible, we will obtain or calculate value-added estimates based on student achievement to measure teacher effectiveness. In addition, if possible, we will also use districts' measures of effectiveness.

Estimating the effectiveness of key program features. We will conduct exploratory analyses to assess whether particular features of DPBIP are associated with impacts on student achievement. These analyses will, in particular, examine the relative effectiveness of DPBIP models that place different weights on individual versus group performance in the determination of incentive payouts. Other programmatic features of interest include the average and maximum size of the incentive payouts and the degree to which the payouts vary across educators.

Since we do not expect that districts will randomly assign specific components of their DPBIP to schools, we will not be able to experimentally assess the relative effectiveness of different DPBIP program features. Instead, we will examine the association between impacts and key program features in a regression framework. We will be careful to note that an observed association between impacts and programmatic features may not necessarily have a causal interpretation.

For these analyses, we will rely on findings from the implementation analysis to examine how the variation in programmatic features is related to the impact. Our basic approach is to regress the estimated district-specific impacts from equation (1) on a measure of a specific programmatic feature. For the estimated impact $\hat{\beta}_k$ from district k, we estimate:

$$(2) \quad \hat{\beta}_k = \pi_0 + \lambda W_k + \omega_k$$

where π_0 is an intercept, W_k is a measure of a specific programmatic feature with associated coefficient λ, and ω_k is an error term that includes random error in estimating the true impact β_k.

Because impacts might be more precisely estimated in some districts than in others, we will weight grantees by the precision of the estimated impacts when estimating equation (2) to account for this source of heteroskedasticity in the error term. For each of the programmatic features described earlier, we will estimate equation (2) with the specified program feature as the only covariate, given the limited number of grantees in the sample.

Understanding the implementation experiences of TIF districts. Understanding the implementation experiences and challenges of TIF grantees will provide essential information for improving the implementation of future incentive programs and is crucial for the interpretation of the impact findings. We will analyze the implementation data collected from grantee, district, and school documents; district, principal, and teacher surveys; and telephone interviews with districts to report on their incentive policies and experiences. Since the evaluation districts were purposively selected, and the impact estimates cannot necessarily be generalized beyond this sample, we will use the district surveys to construct tables on their incentive policies, comparing the evaluation districts to all recent awardees. We also will use the district surveys and information from telephone interviews to document and analyze implementation challenges. The principal and teacher surveys will provide critical context to determine if they understood the incentive compensation policy and program in their district and school and adjusted their behavior accordingly. After the initial survey, for each subsequent wave of the principal and teacher surveys, we will construct tables to assess any changes in educators' understanding and behavior.

Comparing the outcomes for TIF districts to non-TIF districts. In addition to estimating the impact of the DPBIP, we will plan to tabulate outcomes for a group of TIF schools that includes both treatment and control group members, and a reference group of non-TIF schools that are not implementing any kind of PBCS. The goal of this analysis is to provide information on the broader set of TIF-funded reforms beyond performance pay. Outcome data for non-TIF schools, such as average test scores, and PBCS implementation status will be obtained from publicly available data sources.

b. Publication Plans

We will prepare four reports presenting the results of these tabulations. The first report, with a projected release date of November 2013, will describe districts' implementation strategies and challenges and examine first-year impacts. The second, third and fourth reports, scheduled for release in fall 2014, 2015, and 2016, respectively, will present cumulative as well as yearly impacts. Reports will be written in a style and format accessible to policymakers and research-savvy practitioners and will comply fully with the standards set by the National Center for Education Statistics.

17. Approval Not to Display the OMB Expiration Date

The study will display the OMB expiration date.

18. Explanation of Exceptions

No exceptions are being sought.

REFERENCES

Anderman, C., A. Cheadle, S. Curry, P. Diehr, L. Shultz, and E. Wagner. "Selection Bias Related to Parental Consent in School-Based Survey Research." *Evaluation Review*, vol. 19, no. 6, 1995, pp. 663–674.

Constantine, Jill, Daniel Player, Tim Silva, Kristin Hallgren, Mary Grider, and John Deke. "An Evaluation of Teachers Trained Through Different Routes to Certification." Princeton, NJ: Mathematica Policy Research, February 2009.

Decker, Paul, Steven Glazerman, and Daniel Mayer. "The Effects of Teach For America on Students: Findings from a National Evaluation." Princeton, NJ: Mathematica Policy Research, June 9, 2004.

Eaton, Danice K., Richard Lowry, Nancy D. Brener, Jo Anne Grunbaum, and Laura Kann. "Passive Versus Active Parental Permission in School-Based Survey Research: Does the Type of Permission Affect Prevalence Estimates of Risk Behaviors?" *Evaluation Review*, vol. 28, no. 6, 2004, pp. 564–577.

Garet, Michael S., Stephanie Cronen, Marian Eaton, Anja Kurki, Meredith Ludwig, Wehmah Jones, Kazuaki Uekawa, Audrey Falk, Howard Bloom, Fred Doolittle, Pei Zhu, and Laura Sztejnberg. "The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement." Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, September 2008.

Glazerman, Steven, Paul Decker, and Daniel Mayer. "Alternative Routes to Teaching: The Impacts of Teach For America on Student Achievement and Other Outcomes." *Journal of Policy Analysis and Management*, vol. 25, no. 1, 2006, pp. 75–96.

Hanushek, Eric A. "Efficient Estimators for Regressing Regression Coefficients." *American Statistician*, vol. 28, no. 2, 1974, pp. 66–67.

Hanushek, Eric A., and S. Rivkin. "Generalizations About Using Value-Added Measures of Teacher Quality." *American Economics Review*, vol. 100, no. 2, 2010, pp. 267-71.

Rivkin, S., E. Hanushek, and J. Kain. "Teachers, Schools, and Academic Achievement." *Econometrica*, vol. 73, no. 2, 2005, pp. 417–458.

Rockoff, J. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* (AEA Papers and Proceedings), vol. 94, no. 2, 2004, pp. 247–252.

Tuttle, Christina, Steven Glazerman, and Tara Anderson. "ABCTE Teachers in Florida and Their Effect on Student Performance." Washington, DC: Mathematica Policy Research, April 27, 2009.

Yoon, K. S., T. Duncan, S.W.Y. Lee, B. Scarloss, and K. Shapley. "Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2007.

MATHEMATICA
Policy Research, Inc.

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research