An Impact Evaluation of the Teacher Incentive Fund (TIF)

Section A

July 21, 2010

MATHEMATICA
Policy Research, Inc.

An Impact Evaluation of the Teacher Incentive Fund (TIF)

Section A

July 21, 2010

MATHEMATICA
Policy Research, Inc.

CONTENTS

TABLES

PART A: SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This OMB package requests clearance to ensure that grantees' program design and program implementation are consistent with the requirements for a rigorous evaluation of the Teacher Incentive Fund (TIF), and if necessary, recruit grantees for the evaluation. This evaluation will include TIF grantees who are awarded funds from the American Recovery and Reinvestment Act (ARRA) of 2009 and the U.S. Department of Education's (ED) fiscal year 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 teacher and principal (hereafter, educators) mobility and retention. 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 1% across-the-board bonus but will not implement any type of DPBIP throughout 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 1% bonus.

The Notice of Final Priorities (NFP) for the TIF grants, published in the *Federal Register* on May 21, 2010, proposed two competitions for grants that will be awarded in 2010—the Main TIF

---

[1] For this document, DPBIP refers to the differentiated incentive pay portion of a grantee's performance-based compensation system (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.

competition and the TIF Evaluation competition; applicants apply to one or the other competition. Unsuccessful applicants for the evaluation grant will automatically be considered for the Main TIF competition. Successful applicants for the Evaluation competition will receive an "evaluation grant" that includes an additional financial award to fund TIF program activities, including for some uses that are not eligible for funding under the Main competition.[2] Grantees awarded an evaluation grant must demonstrate their ability and agreement to meet the grant requirements, which includes the Main competition requirements plus additional ones specific to the evaluation. Even so, we anticipate that we will need to work with grantees to confirm the requirements of the evaluation and to ensure their successful participation.

This is the first of two requests for the evaluation. A future request will seek clearance to collect educator and student records from districts, administer grantee and educator surveys, and conduct grantee interviews. We are submitting the package in two stages because ensuring that grantees' program design and program implementation are consistent with the requirements of the evaluation must begin before all the data collection instruments are developed and pretested. Also included in this first request is the draft letter to participating districts and principals (Appendix A), an information sheet that will be included with the district/school letter (Appendix B), Mathematica's internal confidentiality pledge (Appendix C), and topics to be discussed and goals of the initial and follow-up phone conversations and site visits that will occur shortly after grants are awarded (Appendix D).

We provide an overview of the study's eventual data collection plans in order to provide context, but they are not the focus of this request. We believe it is also important to note that our eventual data collection plans will differ from those for a study on TIF grantees being conducted by

---

[2] The NFP states an evaluation grantee will receive, minimally, an extra $1 million, and can receive as much as $2 million.

Policy and Program Studies Services (PPSS) in the Office of Planning, Evaluation and Policy Development at ED. First, the two data collection efforts target different respondents. The PPSS study includes grantees from the FY2007 awards while participants in the current study will receive their grants in FY2010. Second, the focus and design of each study is different. The PPSS evaluation is an implementation and feasibility study. Its aim is to describe grantees' program features and implementation experiences, as well as examine the feasibility of using extant data to examine the association between TIF participation and student achievement and educator outcomes. 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. For these reasons, the data collection requirements for this evaluation differ from the current PPSS study.

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 American Recovery and Reinvestment Act of 2009 (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 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 four 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). Research has 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 provide important evidence on how to maximize these projects' effectiveness, 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 teachers and principals may be able to (1) improve student performance in high-need schools 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 teacher compensation reform, if performance-based teacher and principal compensation has an effect on student achievement, and what other aspects of PBCSs are associated with student achievement.

b.   Research Questions

The study's primary research question is:

- What is the impact of DPBIP on student achievement and educator mobility and recruitment?

The study will also address the following secondary research questions:

- Is a particular type of DPBIP model—for example, school- or individual-based or mixed programs—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 primary research question, this study will use an experimental design. Schools within a district will be randomly assigned to either a treatment or control group. Both treatment and control schools will implement the same non-DPBIP components of their project; however, only treatment schools will include a DPBIP component. Control schools will implement all non-differentiated performance-based components of the PBCS *and* provide an across-the-board 1% educator bonus. Control schools will not be permitted to implement a DPBIP component for the duration of the TIF grant.

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 the groups being studied. In the random assignment design, the simple difference between outcomes in treatment and outcomes in control schools is an unbiased estimate of the impact of the grantee's DPBIP component.

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, the study will use non-experimental analyses to address the secondary research questions. To the extent possible, the study will examine the correlation between different types of DPBIP models (individual-, group-based, or mixed) and student and educator outcomes. The ability to separately analyze different DPBIP

models will depend on the number of each type of model implemented by the grantees. Similarly, the study will 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 200 schools and 4,478 teachers. Assuming an average of 10 schools per evaluation grantee, the study will include 20 grantees.[3] The study is designed to detect student achievement gains of .09 of a standard deviation. Though this may be a larger effect than can be obtained in the first year or two of the program[4], 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.

Since grantees must identify the districts and schools that will participate in the evaluation, the sample size will be determined by the selection of grantees awarded a TIF evaluation grant. If the evaluation competition falls short of providing a sufficient number of schools for the study, we will contact and recruit grantees who have been awarded grants from the Main TIF competition to participate in the evaluation.

As part of the evaluation, Mathematica will collect data on educators in treatment and control schools as well as district administrative records to estimate the impact of DPBIP on student achievement and educator mobility and recruitment. We expect that district administrative records

---

[3] For more details, see section B.

[4] For example, Hanushek and Rivkin (2010) summary of studies of teacher effectiveness concludes that having a teacher in the 75th percentile compared to the 25th percentile could cause an increase in student achievement in math of .20 standard deviations.

will allow us to track educator mobility; however, we believe that we may need to rely on data from the educator surveys to supplement educator mobility information for some districts. In addition, the educator surveys will provide critical information on their understanding of the DPBIP program, reasons educators move during the study, recruitment, and background information, such as demographic characteristics and teaching experience. This design, along with our data collection efforts, will allow us to address the key research questions.

d.   Ensuring Grantee, District, and School Successful Participation

Grantee applications will identify the district(s) and schools within the district(s) that will participate in the evaluation. However, we anticipate that we will need to work with grantees, districts, and schools to ensure their planned PBCS and its implementation is consistent with the evaluation requirements. We will also confirm with grantees that they understand and will meet the requirements for the evaluation as outlined in the grant notice. These requirements include the core elements in the main competition and the following additional requirements:

- **Minimum number of schools.** Grantees must identify eight or more schools in tested grades (grades 3–8) that will participate in the evaluation. In addition, there must be a minimum of two schools per school level (for example, if elementary schools are in the evaluation, there must be at least two elementary schools participating).

- **Random assignment.** Grantees must agree to allow Mathematica to randomly assign schools in the evaluation to a treatment or control group.

- **Control schools.** Control schools must implement the same non-differentiated components of the district's PBCS and a 1% across-the-board bonus in place of the DPBIP component; however they may not implement any type of DPBIP component for the duration of the TIF grant.

- **Data requirements.** Grantees, districts, and schools must cooperate with evaluation data collection efforts and provide district records on student achievement and available data on educator's school assignments and background characteristics.

Immediately after the evaluation grants have been awarded, we will begin the process of ensuring grantees successful participation in the evaluation. This includes:

- **Notification letter.** After being notified of their award, we will immediately send a letter (Appendix A) to superintendents and principals in the evaluation. The letter highlights the importance of learning about the effectiveness of incentive programs, reminds the grantee of their participation in the evaluation, and provides an overview of the study

design. The letter will also indicate that a member of the study team will be calling soon to provide more details, discuss the district's participation in the study, and arrange for an in-person meeting with district and school officials.

- **Nontechnical information sheet**. Along with the notification letter, we will send an information sheet (Appendix B). This document describes the random assignment process in a simple and nonthreatening way, includes a partial list of evaluation requirements, delineates the benefits of participation, and presents data collection activities and a timeline. It also identifies the organizations comprising the evaluation team and contact information.

Mailings will be sent via FedEx for quick delivery and to better capture the recipients' attention.

- **Initial and Follow-up calls.** Within a week of sending the notification materials, an evaluation team member will call the grantee to identify the appropriate contact. In subsequent calls, we will briefly describe the study, answer immediate questions, and confirm the district's agreement to participate. We will also arrange for an in-person visit with all key stakeholders in the district (Appendix D).

- **In-person meeting.** We will meet with all key stakeholders during a visit to the district (Appendix D). The meeting may include the grant representative, principals, human resources personnel, and union leadership. The purpose will be to: (1) review the districts' planned PBCS; (2) determine if the PBCS meets the study requirements and if not, develop a plan to work with the district to construct one that will; (3) review the data requirements and determine the districts' infrastructure to measure student achievement gains; (4) discuss the technical assistance (TA) available and develop a plan to provide the necessary TA.

- **Provide implementation support and technical assistance.** Given the information obtained from calls and the in-person meeting, the TA team will work proactively with the district to provide the required TA to ensure that the district develops and implements a program consistent with the goals of TIF and the evaluation. TA will be provided to grantees throughout the grant period, but the team will work with districts to help them develop the tools and expertise needed to independently implement the program by the end of the grant period. TA will be provided by a team of experts who are independent of the evaluation team, but who will coordinate with the evaluation team.

Given Mathematica's substantial experience conducting evaluations, we anticipate some attrition over the study period. To ensure we reach our targeted sample size, we will solidify participation with grantees that could include a total of more than 200 schools. In addition, if the evaluation competition does not result in a sufficient sample of schools, we will recruit from among grantees awarded a grant through the main competition. We will prioritize these grantees based on their ability to meet evaluation requirements. If we need to contact grantees from the main competition, we anticipate contacting two to three times as many grantees as will ultimately be

needed because some may not meet the evaluation criteria. We will follow the same process for contacting grantees from the main competition as we do for evaluation grantees.

e.   Data Collection Plan

This package is not to request OMB clearance for data collection at this time, only clearance to ensure that grantees' program design and program implementation are consistent with the requirements and, if necessary, recruit additional grantees from the Main competition for the TIF evaluation. The study includes several complementary data collection efforts that support answers to the research questions. A brief description of each data source and data collection activity is provided below. The forms for these activities will be developed and submitted in a subsequent clearance package along with estimated burden time for each. For purposes of this exposition, data collection plans assume that all evaluation grantees will participate in a one-year planning period.

- **Grantee Survey.** We will administer three surveys to all TIF grant recipients who receive a new grant in 2010, Evaluation competition grantees and Main competition grantees. The first survey—to be administered in fall 2011—will help us to learn about specific features of the incentive program and to understand the approach grantees used to obtain the buy-in as well as any compromises they had to make. The second, a follow-up survey, will be administered in fall 2012 and explore the grantees' experiences in the first year of implementation (or first and second years if the grantee implements the PBCS in fall 2010), any changes they made in their system, and reasons for the changes. The third survey—to be administered in the fall 2014—will focus on grantees' experiences over the longer period and their plans for sustaining the incentive policies.

- **Educator Survey.** An educator survey will be administered in four waves—spring 2012, 2013, 2014, and 2015—to all principals and a representative sample of teachers in study schools. Data from this survey will help assess educators' knowledge and perceptions of the incentives. In the later waves, we will administer surveys to the same educators even if they have left the school, as well as new principals and teachers in the study schools. By surveying educators over time, the educator survey can provide important information on educators' mobility, retention and recruitment. We will ask about background characteristics only once—in the first wave of the survey and only for newly hired educators in the study schools in latter waves.

- **Teacher and Principal Records.** Detailed information on teachers and principals is necessary to verify educator retention and school assignment. However, the type and reliability of data that districts can provide can vary. Ideally, we will collect data from districts on teacher and principal hiring, movement between schools, and attrition. We also will attempt to obtain information about the start and end dates for each educator's school assignment 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. Although we prefer to receive the data in electronic format, we will use data in whatever form is available. These data will be collected from all evaluation grantee districts in fall of 2012, 2013, 2014, and 2015.

- **Student Records.** To analyze impacts on student achievement, we will collect student records data from the study districts in the summer/fall of 2012, 2013, 2014 and 2015. In addition to test scores, we will collect data such as student age, race/ethnicity, English language proficiency, disability status, eligibility for school lunch programs, and mobility within the district. Information on students' demographic and socioeconomic characteristics and their achievement test scores prior to the study school year will be used both to describe the students in the study and to develop more precise impact estimates. Where possible, we will also request student-teacher linked data in order to estimate teachers' value-added score to better understand mobility of high- and low-performing educators.

- **Grantee Interviews.** For a thorough understanding of each grant program's context, implementation strategy, and challenges, we will conduct in-depth phone interviews with each of the TIF evaluation grant program managers in spring/summer 2012, 2013, and 2015. The interviews will allow a conversational exchange to answer the open-ended questions needed to elicit the descriptive information on topics such as implementation experiences and other ongoing school improvement efforts also funded with ARRA resources.

f.   Study Activities and Timeline

The study is expected to be completed in seven years. Most grantees are expected to use the 2010–2011 school year as a planning year to further develop the core element(s), as described in the grant requirements, and to plan for the implementation of their PBCS in fall 2011. These core elements include:

- A plan to effectively communicate to educators (teachers, administrators, and other school personnel) and the community the components of the PBCS.

- The involvement and support by educators and unions in designing the PBCS (when the union is the designated exclusive representative for the purpose of collective bargaining).

- Rigorous, transparent, and fair evaluation systems for educators that differentiate effectiveness using student growth as a significant factor and classroom evaluations conducted multiple times each school year.

- A data management system that can link student achievement data to educator payroll and human resources systems.

- A plan for ensuring that educators understand the specific measures of educator effectiveness included in the PBCS, and receive professional development enables them to use data generated by these measures to improve their practice.

If a grantee demonstrates in its application that it has these five core elements already in place, it will implement its PBCS program in fall 2010, and the evaluation data collection will begin in spring of 2011. A report describing TIF implementation and presenting the first-year impacts will be prepared in fall 2013. Three additional reports (fall 2014, 2015, and 2016) will estimate yearly impacts and cumulative impacts over the duration of the grants.

Table 1 shows the timing of the major study activities. Since this package is requesting clearance to ensure the successful participation of grantees, only the first activity listed in Table 1 applies to this request. However, to provide an overview of the study, we also show the timeline for other major activities.

Table 1. 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	X	X	X	X	X	X	X	X	X		
Conduct grantee survey		X		X				X			
Conduct educator survey			X		X		X		X		
Collect teacher records data from district				X		X		X		X	
Collect student records data from district				X		X		X		X	
Conduct grantee 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

The main purpose of this evaluation is to estimate the impacts of DPBIP on student achievement and educator mobility and recruitment at high-need schools. Educator quality is a

critical input to student learning, but little is known about how to develop a strong educator workforce. DPBIP is an increasingly important education policy to promote improved instruction; however, little is known about the effectiveness of compensation programs that tie educator pay to student performance.

The findings from this study will provide important evidence for school districts and policymakers on the effectiveness of DPBIP on student achievement and educator outcomes. If possible, this evaluation may also 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 association of other key program aspects of DPBIP models with student and educator outcomes, as well as how districts may overcome common implementation challenges.

Table 2 lists the study's research questions and the data collection to support the answers.

Table 2. Research Questions and Data Collection Method

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

[a]Educator survey will include some unique modules for both teachers and principals.

Study findings will be presented in four reports. The data collected by this evaluation will also be available as restricted-use files, serving as a valuable resource for 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. Therefore, as much information as possible will be gathered from existing data sources, such as TIF grant application packets submitted by awardees (provided by ED) and administrative records (electronic files provided by districts).

The research team will discuss the study design and other logistical details with key staff in the districts through phone calls and site visits to districts. This will also allow us to collect in-depth preliminary information from districts and more efficiently respond to their questions.

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. Although 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 ED evaluations. First, the two data collection efforts target different respondents. The PPSS study includes grantees from the FY2007 awards while participants in the current study will receive their grants in FY2010. In addition, the program requirements differ between the two cohorts (i.e., those studied by PPSS and NCEE). Although it is possible some grantees may overlap, the 2010 grantees would cover new schools or different educators. However, if there are grantees participating in both efforts, we will coordinate with PPSS to avoid requesting duplicate information from grantees.

econd, the focus and design of the studies is different. The PPSS evaluation is an implementation and feasibility study. Its aim is to describe grantees' program features and implementation experiences, as well as examine the feasibility of using extant data to examine the association between TIF participation and student achievement and educator outcomes. 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. For these reasons, the data collection requirements for this evaluation differ from the current PPSS study.

5.   Methods to Minimize Burden on Small Entities

The primary entities for the study are school districts. Burden is minimized for all respondents by requesting only the minimum data required to meet study objectives. Sample sizes and data requirements were also determined by careful consideration of the information needed to meet the study objectives and will be reviewed by the study's technical working group (TWG) before the OMB package for the data collection is submitted.

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 affect student achievement and how grant recipients design, communicate, and implement TIF programs.

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, page 22576 of the *Federal Register* on April 29, 2010. No public comments have been received.

b.   Consultations Outside of the Agency

The study team will work with IES to identify experts in teacher compensation, evaluation methodology, and education policy to become members of the TWG. Once they have been determined, we will seek their input on the evaluation's design.

c. Unresolved Issues

None.

9. Payments or Gifts

The study does not plan to give gifts to districts for participating in the recruitment process.

10. Assurances of Confidentiality

Although the data collection efforts (which will include educator and student records, educator surveys, grantee surveys and interviews) are not the focus of this clearance package and will be the focus of a second clearance request, they will be conducted in accordance with all relevant regulations and requirements. These include the Education Sciences Institute Reform Act of 2002, Title I, Part E, Section 183, that requires "[a]ll collection, maintenance, use, and wide dissemination of data by the Institute … to conform with the requirements of section 552 of Title 5, United States Code, the confidentiality standards of subsections (c) of this section, and sections 444 and 445 of the General Education Provisions Act (20 U.S.C. 1232 g, 1232h)." These citations refer to the Privacy Act, the Family Education Rights and Privacy Act, and the Protection of Pupil Rights Amendment. In addition, for student information, the project director will ensure that all individually identifiable information about students, their academic achievements and families, and information with respect to individual schools shall remain confidential in accordance with section 552a of Title 5, United States Code, the confidentiality standards subsection (c), and sections 444 and 445 of the General Educations Provision Act.

Subsection (c) of Section 183, referenced above, requires the director of IES to "develop and enforce standards designed to protect the confidentiality of persons in the collection, reporting, and publication of data." The study will also adhere to requirements of subsection (d) of Section 183 prohibiting disclosure of individually identifiable information as well as making the publishing or inappropriate communication of individually identifiable information by employees or staff a felony.

Mathematica, and its subcontractors Chesapeake Research Associates and Vanderbilt University, will protect the confidentiality of all information for the study and use it for research purposes only. When reporting the results, data will be presented only in aggregate form, such that individuals and institutions will not be identified. A statement to this effect will be included with all requests for data. All members of the study team with access to the data will be trained and certified on the importance of confidentiality and data security. All data will be kept in secured locations and identifiers will be destroyed as soon as they are no longer required.

The following safeguards are routinely employed by Mathematica to carry out confidentiality assurances during the study:

- All Mathematica employees sign a confidentiality pledge (Appendix C) emphasizing its importance and describing their obligation.
- Identifying information is maintained on separate forms and files, which are linked only by sample identification number.
- Access to hard copy documents is strictly limited. Documents are stored in locked files and cabinets. Discarded materials are shredded.
- Computer data files are protected with passwords and access is limited to specific users.
- Especially sensitive data are maintained on removable storage devices that are kept physically secure when not in use.

11. Additional Justification for Sensitive Questions

We do not anticipate that any of the recruitment data collection will contain items considered to be of a sensitive nature.

12. Estimates of Hours Burden

Table 3 reports the estimated burden hours for 40 district representatives and 200 school principals to participate in initial phone calls and site visits that will occur during the first six months after grants have been awarded. These estimates are based on our experience recruiting schools and districts for evaluation studies. Burden estimates for other data collection efforts described in

section 1.e. "Data Collection Plan" above that are not the subject of this request will be included in a

future request.

Table 3. Estimated Response Time

Respondent	Number of Responses	Unit Response Time (Hours)	Total Response Time (Hours)
District Staff			
Superintendent or grant manager (**20 respondents**)			
Initial phone call	20	1	20
In-person meeting	20	2	40
Follow-up phone call	20	2	40
Other district staff (**20 respondents**)			
In-person meeting	20	2	40
Total district staff	**80**	**7**	**140**
School Principals (200 respondents)			
In-person meeting	200	2	400
Follow-up phone call	200	1	200
Total Principals	**400**	**3**	**600**
Overall Total	**480**		**740**
Annual total	**160**		**247**

The total of 740 burden hours includes an initial one hour phone conversation with the TIF

grant manager or superintendent, a two hour in-person meeting with district staff (the

superintendent or grant manager and one other staff member per district) and school principals, two

hours of follow-up discussions with the TIF grant manager or superintendent, and one hour of

follow up conversations with school principals. The average response rate is 1.67 hours/response

for superintendents/grant managers (5 hours for 3 responses), 2 hours/response for other district

staff, and 1.5 hours/response for principals (3 hours for 2 responses). We estimate that the total

number of respondents will be 240 (20 superintendents/grant managers, 20 district staff, and 200

principals) (or 80/year), the total number of responses will be 480 (160/year) and the total response

time to be 740 hours (247/year).

13. Estimates of Cost Burden to Respondents

There are no start-up costs for respondents.

14. Estimates of Annual Costs to the Federal Government

The total estimated cost of the effort to ensure that grantees' program design and program implementation are consistent with the requirements for the evaluation is $784,000. The cost of the evaluation is $12,000,000 (including the $784,000); the estimated average annual cost of the study over seven years is $1,714,286.

15. Reasons for Program Changes or Adjustments

This is a new data collection resulting in a program change of 247 hours to ensure that grantees' program design and program implementation are consistent with the requirements for a rigorous evaluation.

16. Plan for Tabulation and Publication of Results

We discuss our plans below for tabulating data for all reports to address the research questions and publishing results. Some of the plans are for reference only, as they include data that will be collected after a future request for clearance is approved.

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 implement a DPBIP 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 educator mobility and recruitment. Additional non-experimental 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 educator outcomes, and to learn about grantees' 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 their 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 grantee-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 grantee k:

$$y_{ijk} = \sum_{k=1}^{K} \left(\alpha_k G_k + \beta_{1k}(T_{jk}^{(1)} \times G_k) + \beta_{2k}(T_{jk}^{(2)} \times G_k) \right) + \mathbf{X}_{ijk}\boldsymbol{\delta} + \mathbf{Z}_{jk}\boldsymbol{\gamma} + u_{jk} + \varepsilon_{ijk}$$

(1)

where G_k is a dummy variable for grantee k; α_k is a grantee fixed effect; $T_{jk}^{(1)}$ and $T_{jk}^{(2)}$ are dummy variables for being assigned to a group- or individual-based DPBIP model, respectively; X_{ijk} is a vector of individual baseline characteristics (i.e. if individual, i is a student, X_{ijk} is a vector of student characteristics, and if individual i is a teacher, X_{ijk} is a vector of teacher characteristics); Z_{jk} is a vector of baseline school-level characteristics; u_{jk} is a random school effect; ε_{ijk} is a random individual error term; β_{1k} and β_{2k} are coefficients to be estimated; and δ and γ are coefficient vectors to be estimated. We will estimate equation (1) with ordinary least squares (OLS) using Huber-White ("sandwich")

standard errors that account for school-level clustering. As a robustness check, we will also estimate equation (1) as a hierarchical linear model (HLM).

The coefficients β_{1k} and β_{2k} represent grantee-specific impacts of, respectively, the group-based (or a mixed model if that is more common) and individual-based DPBIP. To estimate the average impact on schools of group-based incentive systems funded by TIF, we will take a weighted average, denoted by $\hat{\beta}_{1\bullet}$, of $\hat{\beta}_{1k}$ across grantees that implement such incentives, where the weights are the number of schools in the grantee subject to that type of incentive. Similarly, a weighted average, $\hat{\beta}_{2\bullet}$, of $\hat{\beta}_{2k}$ provides an estimate for the average impact of individual incentives. Taking a weighted average of estimated impacts across grantees and incentive categories provides an estimate of the average impact of all DPBIP models in the sample. In each case, the standard error of the average impact estimate can be calculated from the estimated variances and covariances among the grantee-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 of 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 2012, 2013, 2014, and 2015. Because the retention outcome is dichotomous, we will estimate the probit model analog of equation (1). School-level teacher data from study schools in fall of 2012, 2013, 2014, and 2015 (from district records) and spring 2012, 2013, 2014, and 2015 (from the educator survey) 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. For the analysis of these school-level composition outcomes, equation (1) can be aggregated to the school level.

To better understand mobility of high- and low-performing educators, for grantees where we can obtain or calculate value-added estimates, we will also estimate a model of transitions that includes a teacher or school value-added measure based on student achievement, and interactions of value added with treatment indicators in the set of independent variables. The coefficients on the value-added by treatment interactions provide an estimate of whether differences in retention between high value-added and lower value-added educators 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 value-added estimates may be endogenous if DPBIP induces greater teacher effort, these estimates are nonexperimental and will need to be interpreted with caution.

Estimating the relative effectiveness of group- and individual-based DPBIP. To estimate the relative effectiveness of a particular DPBIP, we will consider the separate impacts of group-

based or mixed and individual-based incentives. In the ideal design, if districts randomize schools

into two treatment schemes and the control condition, then $(\hat{\beta}_{1k} - \hat{\beta}_{2k})$ is an experimentally based,

unbiased estimate of the impact of group incentives relative to individual incentives in grantee *k*.

Taking a weighted average of $(\hat{\beta}_{1k} - \hat{\beta}_{2k})$ across such grantees (with weights equal to the number of

treated schools) provides an unbiased estimate for the average impact of group incentives relative to

individual incentives. However, we expect that most districts will choose to implement only one

TIF-funded incentive scheme. In this case, within the experimental framework, we will only be able

to examine the effect of an individual- or group-based DPBIP compared to a 1% across-the-board

bonus and no DPBIP component. We will not be able compare individual- and group-based models

to each other. However, a comparison of the magnitude of $\hat{\beta}_{1\bullet}$ and $\hat{\beta}_{2\bullet}$ will provide an exploratory

comparison of group-based or mixed and individual-based incentives on the basis of between-

district variation.

Estimating the effectiveness of other key program features. Since we do not expect that

districts will randomly assign specific components of their DPBIP to schools, we will also not be

able to experimentally assess the relative effectiveness of other features of the DPBIP, such as the

relative weight of student achievement. Instead, we will examine the association between impacts

and key program features in a multivariate regression framework. These exploratory analyses are

designed to shed light on the attributes of DPBIP models that show promise in changing educator

behavior and student achievement. 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 grantee-specific impacts from equation (1) on programmatic features. For the estimated

impact $\hat{\beta}_{ck}$ of incentive category c (group or individual incentives) from grantee k, we estimate:

$$\text{(2)} \quad \hat{\beta}_{ck} = \pi_0 + \mathbf{W}_{ck}\boldsymbol{\lambda} + \omega_{ck}$$

where π_0 is an intercept, W_{ck} is a vector of programmatic features with associated coefficient vector

λ, and ω_{ck} represents random error in estimating the true impact β_{ck}. We will estimate equation (2)

with generalized least squares using the method described by Hanushek (1974).

Understanding the implementation experiences of TIF grantees. 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, grantee and educator surveys, and phone interviews with grantees to report on

their incentive policies and experiences. Since the evaluation grantees were purposively selected, and

the impact estimates cannot necessarily be generalized beyond this sample, we will use the grantee

survey to construct tables on their incentive policies, comparing the evaluation grantees to all recent

awardees. We also will use the grantee surveys and information from phone interviews to document

and analyze implementation challenges. The educator survey will provide critical context to

determine if they understood the incentive compensation policy and program in their district and

school and adjusted their behavior in response. After the initial survey, for each subsequent wave of

the educator survey we will construct tables to assess any changes in educators' understanding and

behavior.

b. Publication Plans

We will prepare four reports presenting the results of these tabulations. The first report, with a

projected release date in November 2013, will describe grantees' implementation strategies and

challenges and examine first-year impacts. The second, third and fourth reports, scheduled for

release in fall 2014, 2015, and 2016, 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 (NCES).

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

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.

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., and Steven G. Rivkin. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economics Review*, vol. 100, no. 2, forthcoming May 2010.

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

Rivkin, Steven, Eric Hanushek, and John Kain. "Teachers, Schools, and Academic Achievement." *Econometrica*, vol. 73, no. 2, 2005, pp. 417–458.

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

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, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L. 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