



AMERICAN INSTITUTES FOR RESEARCH[®]

**Impact Evaluation of
Teacher and Leader Evaluation Systems**

**OMB Clearance Request
For Data Collection Instruments,
Introduction and Part A**

October 19, 2012

**Prepared for:
U.S. Department of Education
Contract No. ED-IES-11-C-0066**

**Prepared by:
American Institutes for Research**

Contents

	Page
Introduction.....	1
Description of the Impact Evaluation of Teacher and Leader Evaluation Systems	2
Purpose.....	2
Research Questions and Theory of Action.....	2
Intervention Selection and Characteristics.....	5
Data Collection.....	8
Supporting Statement for Paperwork Reduction Act Submission.....	13
Justification (Part A).....	13
References.....	23

Exhibits

Exhibit 1. Simplified Theory of Action.....	4
Exhibit 2. Domains and Constructs for FFT and CLASS.....	7
Exhibit 3. Summary of Data Collection Instruments and Schedule.....	9
Exhibit 4. Expert Reviewers.....	16
Exhibit 5. Hour Burden for Respondents.....	20
Exhibit 6. Schedule for Dissemination of Study Results.....	21

Introduction

This request is the second of a two-stage clearance request to carry out data collection activities for the Impact Evaluation of Teacher and Leader Evaluation Systems (TLES) Study. The purpose of this study is to examine the implementation and impacts of a package of evaluation system components that reflects current federal policy.

In May 2012, the Office of Management and Budget (OMB) approved the first clearance request (OMB 1850-0890), which described the study, the design, and the recruitment activities. The recruitment activities included contacting a sample of school districts to establish their eligibility for the study and further contacts and visits with the eligible school districts.

In this second request, the Institute of Education Sciences (IES) of the U.S. Department of Education requests clearance for the study's data collection instruments, specifically the teacher survey, the principal survey, the district interview, and the district archival records collection protocol. Information about other data collections (e.g., intervention fidelity data) related to this study is included in this request to provide a complete picture of the study.

This request contains three major sections.

1. Description of the Impact Evaluation of Teacher and Leader Evaluation Systems
 - Purpose
 - Research questions and theory of action
 - Treatment selection and characteristics
 - Data collection
2. Supporting Statement for the Paperwork Reduction Act Submission
 - Justification (Part A)
 - Description of statistical methods (Part B)
3. Instruments for which we are requesting clearance
 - Teacher survey
 - Principal survey
 - District interview
 - District archival records collection protocol



Description of the Impact Evaluation of Teacher and Leader Evaluation Systems

Purpose

The TLES Study is designed to examine the implementation and impacts of a package of evaluation system components that reflects current federal policy. In contrast to traditional practices, the package of system components provided through the study will use multiple, valid measures, including student growth, to meaningfully differentiate the performance levels of teachers and principals. It will draw on multiple assessments of teacher practice and principal leadership that provide timely feedback to guide the efforts of educators and those who supervise and support them.

To measure impacts on student achievement and other outcomes, a purposive sample of school districts with traditional evaluation systems has been recruited to pilot the study's system. These school districts were selected on the basis of their current evaluation system practices, data infrastructure, and interest in using educator performance information to improve student achievement. A pool of eligible elementary and middle schools was identified in each selected school district and randomly assigned to either treatment or control conditions, with equal numbers of treatment and control schools in each school district. In the treatment schools, the study's teacher and leader evaluation system components were introduced in summer 2012 and will be implemented during the 2012–13 and 2013–14 school years, with support from American Institutes for Research (AIR). In the control schools, the district's current evaluation system practices will continue to be used. During the two-year implementation period, data will be collected to support analyses of the implementation and the impact of the pilot evaluation system, with a final collection of data for impact analyses in fall 2014.

To ensure objectivity, the study will be carried out by two independent teams. The implementation team will design and implement the intervention at the sites recruited for the study, and the evaluation team will develop the study instruments and collect and analyze the data.

Research Questions and Theory of Action

The study has six research questions designed to assess the implementation of the evaluation components provided by the study and its impacts.

- RQ1.** How was the intervention implemented (e.g., district decisions about implementation, the cost of the intervention, the fidelity with which the intervention was delivered, and the participation of key actors), and in what context regarding district policies was it implemented?
- RQ2.** Did the intervention produce a contrast between the treatment and the control schools in teachers' and principals' experiences with performance evaluation systems (e.g., how frequently teachers reported being observed)?

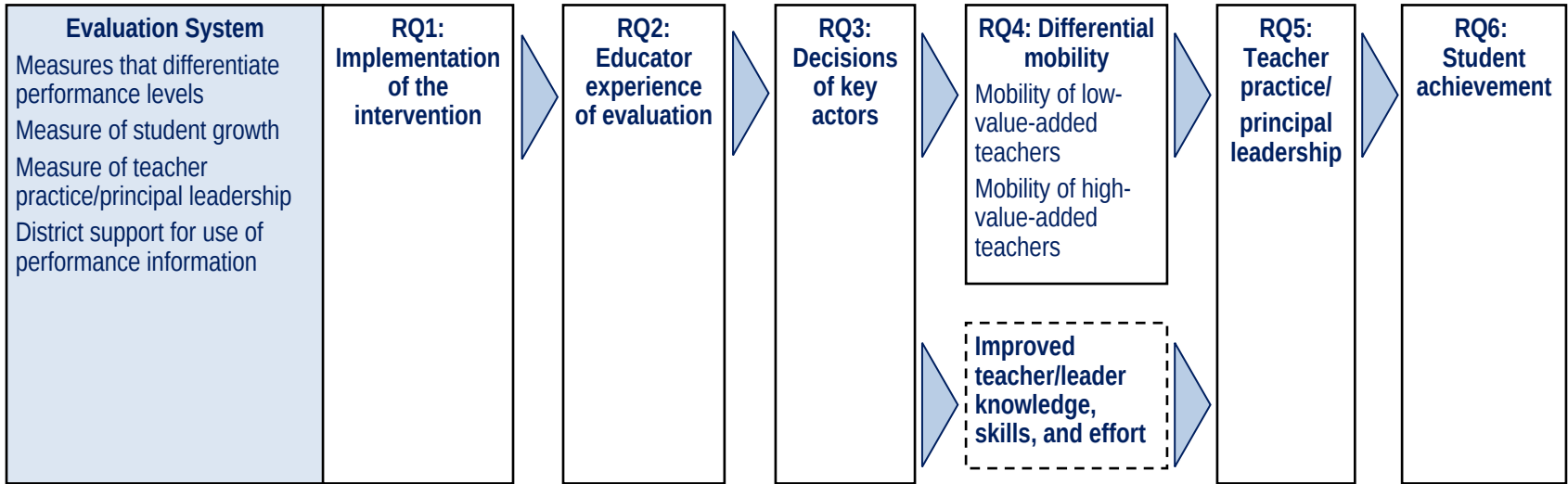
- RQ3.** What impacts did the intervention have on the decisions of key actors (e.g., teachers' decisions to try new techniques, work differently, or pursue learning experiences)?
- RQ4.** What impacts did the intervention have on the mobility of low-value-added teachers and high-value-added teachers?
- RQ5.** What impacts did the intervention have on the dimensions of teacher instructional practice that are the focus of the intervention and on principal instructional leadership?
- RQ6.** What impacts did the intervention have on student achievement?

Because the study intervention is two years in length, we will address the descriptive question (RQ1) in reference to Year 1 and Year 2 separately. We will examine the impact questions (RQ2, RQ3, RQ4, RQ5, and RQ6) separately at the end of Year 1 and Year 2.¹ Impact analyses using data pertaining to the end of Year 2 will measure the cumulative impact of the two years of the intervention.

The theory of action underlying the study's teacher and leader evaluation system components is depicted in simplified form in Exhibit 1. The implementation of the intervention in the treatment schools is expected to cause a contrast between the treatment and the control schools in educators' experiences of the evaluation systems. By clarifying expectations and providing feedback beyond what traditional systems provide, the evaluation system components may influence the decisions of key actors. For example, teachers may decide to try to acquire knowledge of new techniques, refine certain skills, or work differently. The evaluation system components may also influence teacher career mobility, potentially leading low value-added teachers to exit their positions and encouraging high-value-added teachers to stay (i.e., differential mobility). These decisions of key actors are expected to influence teachers' instructional practice and principal leadership, which in turn are hypothesized to improve student achievement.

¹ Because observations of instructional practice will be conducted only during Year 2 of implementation, impact on instructional practice (within RQ5) will be examined only at the end of Year 2.

Exhibit 1. Simplified Theory of Action



Intervention Selection and Characteristics

The study's intervention will consist of teacher evaluation components and leader evaluation components. Measures of student growth will be used for both teacher and leaders. In addition, the teacher evaluation component will measure teachers' instructional practice, and the leader evaluation component will measure principal leadership.

This package of evaluation system components is intended to define performance expectations, repeatedly measure performance, and produce actionable reports. In addition, the study school districts are expected to support the use of performance information. We elaborate on each of these objectives in the following subsections.

Performance Expectations

To define performance expectations, our study treatment will use objective, transparent measures for teacher practice, principal leadership, and student growth. Teacher practice will be measured using observation protocols grounded in frameworks for effective instruction and supported by video libraries that illustrate superior performance. Principal leadership will be measured with a survey-based assessment focused on instructional leadership. Student growth will be measured using value-added modeling of achievement data in reading and mathematics for teachers in Grades 4–8.

Repeated Measurement of Performance

Repeated measurement provides a basis for both formative and summative uses of the performance information. The study's evaluation system components will repeatedly measure teachers' instructional practice (two to four observations per teacher per year), principal leadership (two assessments per principal per year), and student growth (annually, using value-added methods).

Actionable Performance Reports

In-person feedback and online reports will provide performance reports tailored to the needs of the educators being evaluated (i.e., evaluatees), their supervisors, and others designated by the central office who are expected to use the performance information. For evaluatees, performance reports will indicate the individual performance level in the context of clear performance expectations, which will guide their efforts to perform and encourage skill development. Each measurement instance will include opportunities to discuss the measurement results. Observers and assessors will conduct these discussions using protocols that focus discussion on the evaluatee's performance and the performance expectations (Danielson, 2010; Goe, Biggers, & Croft, 2011). For supervisors and others designated by the central office, additional reporting formats will be available to illuminate patterns in the ratings.

District Support for Using the Performance Information

To promote the use of performance information that may improve student achievement, AIR will facilitate planning meetings in each school district to facilitate districts' decisions about how the evaluation system components will be used in the treatment schools in their district, including how this use will be supported. For example,

- School districts may choose to use the performance information to support teachers' efforts to engage in professional development opportunities that are tailored to their needs. To do so, a school district may, for example, review its menu of professional development offerings to assess which activities are aligned to the study's observation frameworks and then facilitate teachers' access to those offerings.
- School districts may choose to use the performance information in their current performance evaluation system, as was done in the schools participating in the Excellence in Teaching Program in Chicago Public Schools. In that program, principals incorporated ratings from a pilot observation measure in teachers' official evaluations by taking advantage of a policy that allowed principals to supplement standard performance criteria with "local school requirements" at the beginning of the school year.

In what follows, we describe the three main components of the planned performance measurement system: a component for feedback on student growth, a component for feedback on instructional practice, and a component for feedback on principal leadership.

1. Component for Feedback on Student Growth

Within the first intervention year and at the end of each of the two intervention years, the study will generate evaluation system reports for teachers, principals, and district leaders that summarize teacher and school performance as measured by student growth. Principals and teachers will receive an online webinar on how to interpret these reports. To ensure coherence with existing school processes, principals will be encouraged to integrate these sessions into existing meetings focused on the interpretation and use of student achievement data.

2. Component for Feedback on Instructional Practice

A number of frameworks for instructional practice have been developed for research purposes, including frameworks for specific content areas. Among these frameworks, two have emerged as most suitable for use in teacher evaluation systems because of their applicability across subjects and grades and because of evidence of their validity and connection to student achievement: the Framework for Teaching (FFT) and the Classroom Assessment Scoring System (CLASS; Goe, Bell, & Little, 2008).

The study districts were divided into two groups, such that teachers in the treatment schools in four school districts will receive feedback on instructional practice using FFT, and those in the treatment schools in the other 5 school districts will do so using CLASS. District assignment to one of the two frameworks was nonrandom; district preferences for one or the other framework were taken into account.

With this design, the main impact findings, pooled across all 9 districts, will pertain to frameworks for instructional practice having the general features called for in recent policy developments, rather than one particular approach. FFT and CLASS have many similarities despite their different origins and theoretical underpinnings (Goe et al., 2008). As Exhibit 2 shows, the frameworks have many similar constructs and measurement approaches, although their rating scales are somewhat different: Items are rated on a four-level scale for FFT but a seven-level scale for CLASS.

Exhibit 2. Domains and Constructs for FFT and CLASS

Framework for Teaching	Classroom Assessment Scoring System (CLASS-Upper Elementary and CLASS-Secondary)
<p>Domain 2: Classroom Environment</p> <ul style="list-style-type: none"> • Creating an Environment of Respect and Rapport • Establishing a Culture for Learning • Managing Classroom Procedures • Managing Student Behavior • Organizing Physical Space <p>Domain 3: Instruction</p> <ul style="list-style-type: none"> • Communicating With Students • Using Questioning and Discussion Techniques • Engaging Students in Learning • Using Assessment in Instruction • Demonstrating Flexibility and Responsiveness 	<p>Domain 1: Emotional Support</p> <ul style="list-style-type: none"> • Positive Climate • Negative Climate • Teacher Sensitivity • Regard for Adolescent Perspectives <p>Domain 2: Classroom Organization</p> <ul style="list-style-type: none"> • Behavior Management • Productivity • Instructional Learning Formats <p>Domain 3: Instructional Support</p> <ul style="list-style-type: none"> • Content Understanding • Analysis and Problem Solving • Quality of Feedback • Instructional Dialogue

Note. FFT includes two other domains that are not amenable to measurement through observation and are not included in the study’s measure of teacher practice: Domain 1, called Planning and Preparation, and Domain 4, called Professional Responsibilities.

The implementation of FFT and CLASS in the study districts will follow the same general parameters. For example, for teachers in Grades 4–8 who are responsible for mathematics or reading/English language arts (ELA) instruction, both frameworks will provide four rounds of observation and feedback, and both will use peer observers for three cycles and principals for a fourth cycle. In addition, both frameworks will provide teachers with video libraries that illustrate superior performance.

3. Component for Feedback on Principal Leadership

To define expectations for principal leadership, we will deploy the Vanderbilt Assessment of Leadership in Education (VAL-ED) in treatment schools in all study districts. VAL-ED is a tool for principal evaluation with established validity and implementation fidelity (Condon & Clifford, 2010). The researchers who developed VAL-ED have published its psychometric properties in peer-reviewed journals and websites (www.valed.com/research.html) and have been continuously improving the tool. VAL-ED also is aligned with national standards for principal leadership (Goldring et al., 2009).

VAL-ED focuses on leadership as it relates to teacher and student learning and thus shares the same broad purpose as the study’s teacher and leader evaluation system. VAL-ED gathers data

on principal behaviors from principals, supervisors, and teachers (an approach called “360-degree assessment”) and provides both criterion-referenced and norm-referenced scores. Thus, principals can understand their performance ratings both in terms of specified criteria and relative to other principals evaluated using VAL-ED in the United States.

Data Collection

An overview of the study’s instruments and the schedule for their use is provided in Exhibit 3. The bolded instruments involve burden and are therefore the basis for seeking clearance. The unbolded instruments are listed to provide a complete picture of the study. The exhibit does not include district screening and recruitment data collection activities because clearance has already been received for these activities.

Exhibit 3. Summary of Data Collection Instruments and Schedule

	Treatment	Control	2012–13				2013–14				2014
			Jul–Sep	Oct–Dec	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec	Jan–Mar	Apr–Jun	Jul–Sep
1. Classroom observations	All Grade 4–8 teachers of mathematics and/or reading/ELA (approx. 1,310 teachers)									End of year	
2. Teacher survey	All K–8 teachers of mathematics and/or reading/ELA (approx. 2,940 teachers)					End of year				End of year	
3. Principal survey	70 principals	70 principals				End of year				End of year	
4. District interviews	9 districts					End of year				End of year	
5. Intervention fidelity data	9 districts		Multiple times throughout the year								
6. District archival record data	9 districts				Student records and rosters of study schools		Archival employee and student records		Rosters of study schools		Archival employee and student records
7. Intervention cost records	9 districts						Cost records				Cost records

1. Classroom Observations

Independent classroom observations will be conducted in addition to the classroom observations that are part of the evaluation system. During only Year 2 of the study in spring 2014, all

Grade 4–8 teachers of mathematics and reading/ELA within all 140 study schools will be video recorded for the duration of one class session up to two times. Data from these observations will be used to conduct analyses of the impact of the evaluation system on teacher practice (RQ5).

To enable the analyses of impact on teacher practice to complement the analyses of impact on student achievement, teachers in the independent sample of observations will be from among the grade levels and subjects for which student impacts will be analyzed (i.e., Grades 4–8 in mathematics and reading/ELA). Because teachers will be video recorded, each observation can be coded using both CLASS and FFT. (See Exhibit 2 for the domains and constructs measured by CLASS and FFT.)

2. Teacher Surveys

Teacher surveys will be administered in spring 2013 and spring 2014 to all K–8 teachers of mathematics and reading/ELA within all the 140 study schools. The survey will be used to determine whether the intervention produced a contrast between the treatment and the control schools, particularly with respect to teachers' experiences with performance evaluation (e.g., the number and the duration of classroom observations through which they received a rating or other form of observation-specific feedback; RQ2). The survey will focus on teachers' decisions related to instructional improvement and mobility, as well as their attitudes and beliefs about their ratings and their capacity to change practice in the desired ways (RQ3). Finally, the survey will be used to address whether the evaluation system had an impact on principal leadership practices (RQ5).

3. Principal Surveys

Principal surveys will be administered in spring 2013 and spring 2014 to the principal in each study school. The survey will be used to capture principal time which is an important component of calculating intervention costs (e.g., observing teachers, preparing written feedback, and providing individual feedback during meetings; RQ1). The survey will be used to determine whether the intervention produced a contrast between the treatment and the control schools, particularly with respect to principals' experiences with performance evaluation (e.g., number of times they received a rating of some type for their performance as a principal by a supervisor in the current school year and how those evaluations were conducted) and whether the intervention produced different types of teacher performance information for principals (RQ2). The survey will focus on principals' decisions related to improvement in instructional leadership (e.g., providing varying levels of support to teachers of differing performance levels) and mobility, as well as their attitudes and beliefs about their ratings and their capacity to change practice in the desired ways (RQ3).

4. District Interviews

To answer RQ1, interviews will be conducted in spring 2013 and spring 2014 with the officials in each school district who are responsible for teacher and leader evaluation systems. The interviews will be used to collect information regarding the extent to which the school districts followed through on their plans to implement the intervention, as well as barriers they encountered to implementing their plans and how they overcame those barriers. Information will be collected regarding the integration of the study's evaluation system with existing district processes. The interviews will also be used to collect contextual information regarding the districts' human resources policies in the control schools during the study, focusing on their teacher and principal evaluation system policies and the ways in which performance data are used.

5. Intervention Fidelity Data

Various intervention fidelity records will be collected to answer RQ1. For instance, we plan to capture the fidelity of delivery and participation in key intervention events through in-person visits and the monitoring of online webinars (i.e., collecting of attendance sheets and agendas/schedules). For each district-selected observer and principal, we will obtain information about their credentials, prior experience, and certification test results from the providers' (Teachstone/Teachscape) training materials and online system. Through the providers' online tools, we will capture performance information as well as administrative records. For each observation session and feedback session, the system will provide the session dates and the participant list. Additionally, in the course of completing logs to track their work, the observers will record the focus of the feedback session and use of the video library during the feedback session. Online system records from the VAL-ED provider (Discovery Education) will provide performance information as well as administrative records regarding the number of teachers and district staff who were asked to complete VAL-ED assessments, the VAL-ED survey response rates, the dates when principals received their assessment results, and the dates when principal feedback sessions occurred. Finally, AIR's online system will report value-added scores for all teachers and principals in the treatment schools.

6. District Archival Record Data Collection

Finally, various district archival records will be collected. A brief description of each follows.

Student Records for Value-Added Modeling and Impact Analyses. Student assessment and demographic records will be collected from districts. These data will be used to conduct value-added modeling for the evaluation system reports noted above (see "Component for Feedback on Student Growth" section). These data also will be used to conduct impact analyses (RQ6) for the study reports. To complete the evaluation system reports, the implementation team will request data from the 2008–09 through the 2013–14 school years; to complete the study reports, the evaluation team will require data from the 2011–12 through the 2013–14 school years.

Teacher and Principal Records for Survey Administration and Mobility Analyses. We will request administrative data (e.g., e-mail addresses and grade/subject assignments) for teacher survey administration in January 2013 and January 2014 for each school included in the study.

(As noted earlier, the surveys will be used to answer RQ1, RQ2, RQ3, and RQ5.) In addition, we will request administrative data (e.g., school entry/exit dates and demographic characteristics) for all teachers and principals in the participating school districts for analyses to track mobility between the 2011–12 school year and the 2014–15 school year (RQ4).² For these teachers and principals, we will request teacher placement information for any teacher who worked in a participating school at any time during this time period. These requests will be made twice. In fall 2013 (for the Year 1 report), we will request data corresponding to final, end-of-year 2011–12 records (prior to randomization) and final rosters as they correspond to the start of the 2012–13 and 2013–14 school years. In fall 2014 (for the Year 2 report), we will request final rosters as they correspond to the start of the 2014–15 school year.

Performance Evaluation System Ratings. In summer 2013 and summer 2014, AIR will request the performance ratings given to teachers under the local evaluation system to answer RQ3. In summer 2013, we will request local ratings corresponding to 2011–12 and 2012–13. In summer 2014, we will request local ratings corresponding to 2013–14. We will collect these ratings with linkages to schools rather than teachers; we will request for each school a list of the local ratings of the teachers in the study. These data will be used to answer RQ3, to estimate the impact of the intervention on local ratings given that principals may be more willing to give low ratings in their local systems.

7. Intervention Cost Records

Within RQ1, one aspect of the intervention to be examined is its cost. AIR will use its internal project-cost tracking system to compute the costs of each component of the intervention (i.e., the component for feedback on student growth, the component for feedback on instructional practice, and the component for feedback on principal leadership). Using this system, we will combine AIR assessment costs (i.e., to produce value-added reports), system providers' costs (Discovery, Teachscape, and Teachstone; e.g., to provide training and support to districts), and study costs to conduct observations of teacher instructional practice and to provide feedback.

² These data will provide the basis for estimating impacts on teacher mobility (RQ4) and providing descriptive information on mobility as context for interpreting the study results.

Supporting Statement for Paperwork Reduction Act Submission

A. Justification

1. Circumstances Making Collection of Information Necessary

Improving student achievement is a core priority for federal education policy. In 2011, only 34 percent of fourth graders and 34 percent of eighth graders performed at or above proficiency level in reading (National Center for Education Statistics, 2010b). Similarly, 40 percent of fourth graders and 35 percent of eighth graders attained the proficient level in mathematics (National Center for Education Statistics, 2010a). The performance of U.S. students on international assessments lags far behind that of their international peers (Aud et al., 2011; Hanushek, Peterson, & Woessmann, 2010).

In an effort to improve student outcomes, federal education policy—as manifested in the 2002 reauthorization of the Elementary and Secondary Education Act (ESEA), the American Reinvestment and Recovery Act (ARRA) of 2009, and the ESEA Flexibility waiver program—targets teacher quality as a potential lever for improving student achievement. Several studies support this focus, with recent work examining individual teachers for several years to estimate the degree of variation in achievement gains that can be attributed to teachers rather than measurement error (Bill & Melinda Gates Foundation, 2011; Schochet & Chiang, 2010). On balance, these studies suggest that being assigned to a teacher who is 1 standard deviation above average in effectiveness at raising student achievement may lead to a 0.1 standard deviation increase in student achievement. If such a teacher effect accumulates as students move from one grade to the next, it would translate into a substantial difference in achievement by the time students leave high school. Similar studies suggest that some variation in achievement gains can be attributed to principals, which indicates that principal quality may be another lever for education policymakers (Hallinger & Heck, 1998; Leithwood, Harris, & Strauss, 2010; Leithwood & Jantzi, 2005; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Waters, Marzano, & McNulty, 2003).

The federal government has invested considerable financial resources into supporting efforts to develop systems to measure educator effectiveness and use that information in human resource policies. The Teacher Incentive Fund (TIF) was established in 2006 with an initial funding base of approximately \$100 million to identify and reward effective educators in high-poverty schools, based in part on student growth. The goal of the program is to improve student achievement by increasing the effectiveness of educators in high-poverty schools.

ARRA provided additional funding for TIF and established competitive grants to help states build their pools of effective teachers and principals through the Race to the Top (RTT) program, in which a core priority area is the effectiveness of teachers and leaders. Using RTT grants, states and districts are developing new teacher and leader evaluation systems, with the goal of improving teacher effectiveness and student outcomes. ARRA included \$4.35 billion in funds for RTT. The U.S. Department of Education's (ED's) ESEA Flexibility program provides a

continued emphasis on the development of teacher and leader evaluation systems. Through the Flexibility program, ED relieves states from selected ESEA requirements in exchange for compliance with new requirements, including the implementation of reformed teacher and leader evaluation systems.

Although states and districts across the country have been developing and implementing new evaluation systems to identify and reward highly effective educators, no large-scale study has yet been carried out to confirm the efficacy of these new systems. The TLES Study will address this need. This study is funded by TIF.

This is the second of a two-stage clearance request to carry out data collection activities for the TLES Study. In May 2012, OMB approved the first clearance request (OMB 1850-0890), which described the study, the design, and the recruitment activities. The recruitment activities included contacting a sample of districts to establish their eligibility for the study and visiting eligible districts. In this second request, the Institute of Education Sciences (IES) of the U.S. Department of Education requests clearance for the study's data collection instruments, specifically the teacher survey, the principal survey, the district interview, and the district archival records collection protocol.

2. Purposes and Uses of the Data

The purpose of the TLES Study is to test a package of teacher and leader evaluation system components that is consistent with current federal policy. In contrast to traditional systems, the evaluation system components provided by the study will use multiple, valid measures, including measures of student growth, teacher practice, and principal leadership, to meaningfully differentiate the performance levels. This study will be the first large-scale randomized field trial addressing such evaluation systems. Using an experimental design, the study will examine both implementation and impacts (e.g., impacts on teacher mobility, impacts on student achievement). Data collected by the TLES Study will be of immediate interest and import for policymakers, researchers, and practitioners.

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, and technology will be used to reduce burden for all the data collections for which we are seeking clearance.

The teacher and principal surveys will be administered online using a Web platform that will allow respondents to complete the survey at a time and place that is convenient for each respondent.

District interviews will be conducted via telephone to reduce the writing burden for district personnel. This mode of data collection is appropriate for the conversational exchange necessary to obtain answers to the open-ended questions and allows probing for more detail than a self-administered survey can provide.

For the archival data, we will reduce burden by gathering the data electronically rather than in hard copy. We will provide clear instructions on the data requested and methods of transmitting the data securely.

A toll-free number and an e-mail address will be available during the data collection process to permit respondents to contact AIR with questions or requests for assistance. The toll-free number and the e-mail address will be included in all communication with respondents.

4. Efforts to Identify Duplication

Throughout the evaluation, efforts will be made to minimize and reduce the burden on respondents. Wherever possible, we rely on secondary data sources to reduce burden on district and school personnel. By collecting archival records on teachers and principals (e.g., salary and mobility information), we will eliminate the need for several items on the teacher and principal surveys and thereby avoid redundancy.

In addition, we will make sure to coordinate efforts with other studies going on in the field. Currently, there is an IES impact evaluation of the TIF program that seeks to measure the impact of performance-based compensation on student achievement and the mobility and retention of teachers and principals. The national TIF evaluation measures the impact of performance-based compensation specifically, while our purpose is to test a teacher and leader evaluation system, making the focus of the two studies very different.

5. Methods to Minimize Burden on Small Entities

To be considered a small entity by OMB, a school district would need to have a population of fewer than 50,000 students. Our first step in minimizing burden to small entities was to exclude some of the smallest school districts that might have been the most burdened by the study requirements with certain recruitment criteria. Specifically, our criteria required that districts have (1) at least 10 qualifying schools, (2) data systems that support value-added modeling, and (3) not announced a recent implementation (or planned implementation) of a new teacher or leader evaluation system to take effect before the 2014–15 school year. Second, in working with all districts, we seek to minimize burden to what is necessary. All participating districts need to participate in the study data collections to provide the data necessary for the study's analyses. Thus, we cannot reduce burden on small entities by changing the collections. To avoid placing extra burden on small entities associated with travel, the implementation events are conducted separately in each district. The study team will work closely with districts to ensure the most efficient processes are established for data collection to minimize burden on the part of district personnel.

6. Consequences of Not Collecting the Data

States and school districts are increasingly reshaping teacher and leader evaluation systems, in part due to the federal initiatives described above (see A.1). Without this study, states and districts will have a limited understanding of how these systems affect their students and educators because there have been no rigorous, large-scale studies of the impact of such systems.

7. Special Circumstances

No special circumstances apply to this study.

8. Federal Register Comments and Persons Consulted Outside the Agency

The 60-day notice to solicit public comments was published in Volume 77, Number 151, page 46748 of the Federal Register on August 6, 2012. No comments were received from the public.

To assist with the development of the study, project staff will draw on the experience and the expertise of a network of outside experts who will serve as our technical working group (TWG) members. The consultants and their affiliations are listed in Exhibit 4.

Exhibit 4. Expert Reviewers

Proposed Expert Reviewer	Professional Affiliation
Laura Goe	Educational Testing Service
Thomas Dee	University of Virginia
Daniel McCaffrey	Rand Corporation
Catherine McClellan	Clowder Consulting
Jonah Rockoff	Columbia University
Patrick Schuermann	Vanderbilt University
Carla Stevens	Houston Independent School District
John Tyler	Brown University
Judy Wurtzel	Independent consultant

To date, the project advisor and the TWG members have convened twice (once in person and once via teleconference) to discuss the study design and data collection. Project staff will also consult outside experts individually on an as-needed basis.

9. Payment or Gifts

Incentives have been proposed for the teacher survey to partially offset respondents' time and effort in completing the survey. We propose offering a \$25 incentive to a teacher each time he or she completes a questionnaire to acknowledge the 30 minutes required to complete each questionnaire.

Incentives are also proposed because high response rates are needed to make the survey findings reliable, and we are aware that teachers are the targets of numerous requests to complete surveys on a wide variety of topics from state and district offices, independent researchers, and ED. Although some districts will have solicited buy-in from teachers to participate in the evaluation, our recent experience with teacher surveys supports our view that monetary incentives are needed to ensure adequate response rates.

No incentives will be given to principals for completing the principal survey or to district staff for completing the district interview or the collection of district archival records.

10. Assurances of Confidentiality

A consistent and cautious approach will be taken to protect all information collected during data collection. This approach will be in accordance with all relevant regulations and requirements. These include the Education Sciences Institute Reform Act of 2002, Title I, Part E, Section 183, which 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 Education Provisions Act.

Subsection (c) of Section 183 requires the IES director 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 the disclosure of individually identifiable information as well as making the publishing or inappropriate communication of individually identifiable information by employees or staff a felony.

Except with respect to the information provided to educators and their supervisors as part of the study’s package of teacher and leader evaluation system components, AIR will use all information collected in the study for research purposes only. When reporting the study results, data will be presented in aggregate form only, such that individuals and institutions will not be identified. A statement to this effect will be included with all requests for study data. All members of the study team with access to the study data will be trained and certified on the importance of privacy and data security. All study data will be kept in secured locations, and identifiers will be destroyed as soon as they are no longer required.

We will request unique identifiers (IDs) for each student, as well IDs for the mathematics and reading/ELA teacher(s) associated with each student. To minimize access to personally identifiable information, we will request that districts send a separate crosswalk file that links teacher names to IDs. Both the implementation and the evaluation teams will maintain data security by allowing access to shared folders only to team members who need it. We also will create a process whereby the project director must approve all requests for access to these folders (and maintain a list of who has access).

To enable us to conduct analyses that link individual value-added scores to both survey responses and classroom observation scores, the implementation team will attach scrambled

teacher IDs to the survey and observation data before the data are made available to the evaluation team.

In addition to these safeguards, AIR routinely employs the following to carry out privacy assurances with respect to study data:

- All AIR employees sign a privacy pledge 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. Justification of Sensitive Questions

No questions of a sensitive nature will be included in the teacher and principal surveys and the district interview. For the district archival records collection protocol, we will request the performance ratings given to teachers under the local evaluation system. In summer 2013, we will request local ratings corresponding to 2011–12 and 2012–13. In summer 2014, we will request local ratings corresponding to 2013–14. We will collect these ratings with linkages to schools rather than teachers; we will request for each school a list of the local ratings of the teachers in the study. These data will be used to estimate the impact of the intervention on local ratings. Whether or not the intervention has an impact on local ratings, and in what direction, is an important aspect of RQ3. In response to the intervention, principals may be more willing to give low ratings in their local systems. Currently, local rating systems tend to produce ratings clustered in the highest performance category (Weisberg, Sexton, Mulhern, & Keeling, 2009).

12. Estimates of Hour Burden

The total estimated hour burden for the data collections for the TLES Study is 4,265 hours. Based on average hourly wages for participants (which were calculated based on estimates of yearly salaries), this amounts to an estimated monetary cost of \$124,285. Exhibit 5 summarizes the estimates of respondent burden for the study activities. The burden estimate for the teacher survey includes time for 85 percent of a sample of 2,940 teachers (treatment and control) in the 9 districts to respond to a 30-minute survey in the spring in each year of the study. The burden estimate for the principal survey includes time for 85 percent of all 140 principals (treatment and control) in the 9 districts to respond to a 30-minute survey in the spring in each year of the study. The burden estimate for the district interview includes time for 100 percent of all 9 district contacts to respond to a 90-minute interview in the spring in each year of the study. The district archival records requests require an estimated 20 hours of burden for 1 district data person to pull the requested data in each district. In Year 1 of implementation, there will be four archival records request: student records from 2008–09

through 2011–12, student records from 2012–13, teacher and principal records for sampling purposes, and educator performance evaluation ratings. In Year 2 of implementation, we will have the same request as in Year 1 (with one fewer student records requests) and will add an additional request in the fall of Year 2 for teacher and principal records for the purposes of tracking retention and mobility in our study schools. Finally, in the fall of the 2014–15 school year, we will be collecting the final round of teacher and principal records for the purposes of tracking retention and mobility in our study schools.

Exhibit 5. Hour Burden for Respondents

	Data Collection	Total Sample Size	Estimated Response Rate	Number of Respondents	Number of Administrations	Total Number of Responses	Time Estimate (in hours)	Total Hours	Hourly Rate	Estimated Monetary Cost of Burden
School Year 2012–13	Teacher survey, spring	2,940	85%	2,499	1	2,499	0.5	1,250	\$25	\$31,238
	Principal survey, spring	140	85%	119	1	119	0.5	60	\$35	\$2,083
	District interview, spring	9	100%	9	1	9	1.5	14	\$35	\$473
	District archival records collection	9	100%	9	4	36	20	720	\$35	\$25,200
	Subtotal							2,043		\$58,993
School Year 2013–14	Teacher survey, spring	2,940	85%	2,499	1	2,499	0.5	1,250	\$25	\$31,238
	Principal survey, spring	140	85%	119	1	119	0.5	60	\$35	\$2,083
	District interview, spring	9	100%	9	1	9	1.5	14	\$35	\$473
	District archival records collection	9	100%	9	4	36	20	720	\$35	\$25,200
	Subtotal							2,043		\$58,993
School Year 2014–15	District archival records collection	9	100%	9	1	9	20	180	\$35	\$6,300
	Subtotal							180		\$6,300
Total across all school years								4,265		\$124,285

13. Estimate of Cost Burden to Respondents

There are no additional respondent costs associated with the data collections for this study other than the hour burden accounted for in item 12.

14. Estimate of Annual Cost to the Federal Government

The estimated cost for all aspects of the study is \$16,783,022 over 5 years, with an average annual cost to the federal government of \$3,356,604. The total Year 1 cost is \$4,668,189; the Year 2 cost is \$5,998,265; the Year 3 cost is \$5,224,346; the Year 4 cost is \$765,151; and the Year 5 cost is \$127,071.

15. Program Changes or Adjustments

This is the addendum package to a clearance request that was approved in May, 2012, to conduct recruitment activities (see OMB 1840-0890). Completion of the data collection activities described in this addendum package requires additional burden totaling 3,374 hours.

16. Plans for Tabulation and Publication of Results

AIR will report the findings from the TLES Study to IES in two substantive reports. The timeline for the dissemination of the study results is summarized in Exhibit 6.

Exhibit 6. Schedule for Dissemination of Study Results

Deliverable	Anticipated Release Date
First report	April 2015
Final report	July 2016

The first report will present the results of the analyses conducted on the data collected for the 2012–13 school year, including mobility as measured in fall 2013. This report will include a description of the study design and the results from both descriptive analyses and impact analyses. More specifically, the descriptive analyses will include the following:

- Descriptive analyses of the characteristics of the study sample
- Analyses of the equivalence of the treatment and the control groups in their background characteristics before intervention implementation
- Descriptive analyses of how the intervention was implemented (e.g., district decisions about implementation, the cost of the intervention, the fidelity with which the intervention was delivered, and the participation of key actors) and in what district policy context it was implemented (RQ1)

In addition, the report will provide results on service contrast (RQ2) and the impacts of the intervention on educators and students. Specifically, the impact analyses will include the following during the first treatment year (2012–13):

- Intermediate outcomes: decisions of key actors (e.g., teachers' decisions to try new

techniques, work differently, or pursue learning experiences; RQ3)

- Outcomes: mobility of teachers and principals (RQ4); teacher instructional practice and principal instructional leadership (RQ5); and student achievement (RQ6)

These analyses will be carried out using hierarchical linear modeling where appropriate to take into account nesting (e.g., the nesting of students and teachers within schools) and will incorporate covariates measured at baseline to maximize precision. To avoid potential selection bias, the impact analyses will employ an “intent-to-treat” approach, in which all students, teachers, and principals in all randomly assigned schools during the 2012–13 school year are included in the analyses, whether or not the teachers or principals actually participated in the intended teacher and leader evaluation system or participated to the full extent expected. This approach is explained further in Part B.

The final report will be a capstone report, focusing mainly on the cumulative impacts of providing the intervention for two years (i.e., 2012–13 and 2013–14). The analyses will parallel those in the first report and will also have impact results for teacher instructional practice. Additionally, using AIR’s internal project-cost tracking system, we will document the costs of each component of the intervention: the component for feedback on student growth, the component for feedback on instructional practice, and the component for feedback on principal leadership. However, the AIR project-cost tracking system will not capture hours spent by school principals; therefore, we will compute principal costs using items from the principal survey.

17. Approval to Not Display OMB Expiration Date

Approval is not being requested; all data collection instruments will include the OMB expiration date.

18. Explanation of Exceptions

No exceptions are requested.

References

- Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., et al. (2011). *The condition of education 2011* (NCES 2011-033). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/programs/coe/pdf/coe_msl.pdf
- Bill & Melinda Gates Foundation. (2011). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: Author. Retrieved from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf
- Condon, C., & Clifford, M. (2010). *Measuring principal performance: How rigorous are commonly used principal performance assessment instruments?* Naperville, IL: Learning Point Associates.
- Danielson, C. (2010). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35–39.
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Goe, L., Biggers, K., & Croft, A. (2011, May). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning*. Presentation at the National Comprehensive Center for Teacher Quality Conference, Washington, DC.
- Goldring, E., Carvens, X., Murphy, J., Porter, A., Elliott, S., & Carson, B. (2009). The evaluation of principals: What and how do states and urban districts assess leadership? *Elementary School Journal*, 110(1), 19–39.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, 9(2), 157–191.
- Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2010). *U.S. math performance in global perspective: How well does each state do at producing high-achieving students?* (PEPG Report No. 10-19). Cambridge, MA: Harvard Kennedy School, Taubman Center for State and Local Government, Program on Education Policy and Governance & Education Next. Retrieved from http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf
- Leithwood, K., Harris, A., & Strauss, T. (2010). *Leading school turnarounds: How successful leaders transform low-performing public schools*. San Francisco: Jossey-Bass.
- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, 4(3), 177–199.

- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. St. Paul: University of Minnesota, Center for Applied Research and Educational Improvement.
- National Center for Education Statistics. (2010a). *The nation's report card: Mathematics 2011* (NCES 2012–457). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/main2011/2012457.pdf>
- National Center for Education Statistics. (2010b). *The nation's report card: Reading 2011* (NCES 2012–458). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/main2011/2012458.pdf>
- Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains* (NCEE 2010-4004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from <http://eric.ed.gov/PDFS/ED511026.pdf>
- U.S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. Washington, DC: Author. Retrieved from <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>
- Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning. Retrieved from http://www.mcrel.org/pdf/LeadershipOrganizationDevelopment/5031RR_BalancedLeadership.pdf
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. Retrieved from <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>