



AMERICAN INSTITUTES FOR RESEARCH®

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

**OMB Clearance Request, Part B**

**February 17, 2012**

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

**Prepared by:  
American Institutes for Research**

## Contents

	<b>Page</b>
B. Description of Statistical Methods.....	1
1. Respondent Universe and Sampling Methods.....	1
2. Procedures for Data Collection.....	5
3. Procedures to Maximize Response Rates.....	5
4. Pilot-Testing Instruments.....	6
5. Names of Statistical and Methodological Consultants and Data Collectors.....	6

## List of Appendices

### Appendix A. Screening Material Documents

- Advance Letter
- District-Level Screening Protocol

### Appendix B. Recruitment Material Documents

- Brief Study Description
- Credentials of the Research Team

## List of Exhibits

Exhibit 1. Mapping of District-Level Screening Protocol Items to Constructs

## ***B. Description of Statistical Methods***

The proposed study focuses on the implementation and impacts the study's teacher and leader evaluation system. To conduct the study, we will randomly assign 12 to 14 schools to two groups in each of the 14 participating districts: one group that will continue using the district's current teacher and leader evaluation system; and another group that will introduce the proposed teacher and leader evaluation system that includes feedback on instructional practice, principal leadership, and student growth. We will collect outcome data from both groups. The average outcome levels in the group of schools not receiving the treatment represents a reliable estimate of the outcome levels that would have been observed in the absence of the treatment. Therefore, the difference in the average outcomes between the treatment schools and the control schools within the same district represents a reliable estimate of the treatment's impact.

This approach is known as the "intent-to-treat" approach, in which all members of the treatment and control groups are included in the impact analysis regardless of their actual participation in the treatment. Following this approach, we will assess the effects of the proposed teacher and leader evaluation system on student achievement by comparing the treatment and control schools in the average reading and mathematics achievement, regardless of the extent to which teachers at each set of schools actually participate in the teacher and leader evaluation system activities associated with the treatment. The effects of the treatment will be estimated separately for each district and then pooled across districts to create an average effect of the treatments (meta-analysis). The resulting intent-to-treat estimates can be interpreted as the effects of being assigned to the given teacher and leader evaluation system rather than the effect of participating in activities provided by the evaluation system. In some respects, these estimates mirror those likely to be observed in real-world settings, where teacher and leader evaluation systems are being used.

In the remainder of Part B, we address the following with respect to study recruitment: respondent universe and sampling, procedures for data collection, procedures to maximize response rates, pilot-testing instruments, and names of statistical and methodological consultants and data collectors.

### ***1. Respondent Universe and Sampling Methods***

The TLES study will test the effectiveness of an approach to teacher and leader evaluation. It will not employ random sampling of districts or schools for the purpose of generalization. Instead, districts will be screened and recruited based on characteristics required by the study design, such as the number of elementary and middle schools in the district that are not currently using a teacher and leader evaluation system similar to the one proposed in the study. To achieve a study sample of 14 districts, AIR will administer screening interviews across a pool of 120 districts, attempt to recruit 30 eligible districts, and establish agreements with 14 districts. Each recruited district will allow the study team to randomly assign a subset of its elementary and middle schools to the two conditions: treatment and control (business-as-usual). The remainder of this section describes our processes for (1) identifying the pool of districts to be screened,

- (2) conducting district-level screening interviews, (3) prioritizing districts for recruitment, (4) recruiting eligible districts, and (5) negotiating final agreements.

### **Identifying the Pool of Districts to Be Screened**

A review of the 2009-10 CCD data as well as district and state website data about teacher and leader evaluation systems will assist us in identifying 120 districts that have the following characteristics:

1. The district operates at least 6 elementary and 4 middle schools.
2. The district is not required by the state to implement a new teacher or leader evaluation system in all schools before 2014–15.
3. The district has not posted anything on its website about recent implementation of a new teacher or leader evaluation system or planned implementation of such a system before 2014–15.

### **Conducting District-Level Screening Interviews**

An informational e-mail will be sent to each of the 120 districts identified using recent CCD data and other data collected about the district’s teacher and leader evaluation system (see Appendix A-1 for the text of this email). This e-mail will include notification that the district meets the initial eligibility requirements for participation and will include documents that briefly describe the study and the team conducting the study (see Appendix B-1 and B-2 for these documents). After the e-mail is sent, an evaluation team member will call each district to inform them about the study and ask them to participate in a telephone interview. The district-level screening protocol that will be used to guide this interview is presented in Appendix A-2 and is described below under the section Procedures for Data Collection.

A district will be determined to be eligible for the study based on a judgment that is made in considering the following additional criteria:

1. The district has the capacity to link student achievement scores to specific teachers in mathematics and reading.
2. The district is not currently using the components of the proposed teacher and leader evaluation system—Classroom Assessment Scoring System (CLASS), the Framework for Teaching (FFT) from the Danielson system, or the Vanderbilt Assessment of Leadership (VAL-ED).
3. There is sufficient contrast between the district’s current teacher and leader evaluation system and the study’s system. The district should not be using an approach that incorporates performance expectations, repeated measurement of performance, and actionable performance reports in a manner that is similar to what is provided in the study’s evaluation system.

The screening protocol is designed to allow early termination of the interview if the district has already implemented a similar system. We anticipate that approximately 45 of the 120 districts will meet the aforementioned criteria.

### **Prioritizing Districts for Recruitment**

Among the 45 districts that we expect will pass the initial screen, some will be more appropriate candidates for the study than others. AIR staff will use the additional information gathered in the screening interviews to prioritize eligible districts for recruitment efforts. The following criteria will be used:

- **Interest.** Districts that signal greater interest will be given higher priority. Interviewers usually receive some signals about the district's level of interest in the study even though the screening protocol contains no questions about interest.
- **Feasibility of implementation.** Districts that have fewer competing initiatives will be given higher priority. In addition, districts with teachers unions that are supportive of the proposed evaluation system may be given higher priority.
- **Geographic diversity.** Some districts may receive higher priority in order to ensure geographic diversity. Although not essential, geographic diversity among the districts would add to the policy relevance of the findings.

Considering these criteria while canvassing the 45 eligible districts will allow AIR to prioritize the districts before initiating requests for site visits.

### **Recruiting Eligible Districts**

Past experience indicates that site visits to districts are necessary to ensure eligibility and reach final agreement on participation. To initiate recruitment efforts, the evaluation team will send materials about the study (see Appendix B-1 and B-2) to officials in eligible districts, beginning with those determined to be of the highest priority. Senior evaluation team members will follow up with these districts by telephone to:

- Determine which district officials must be involved in making the decision about participation
- Communicate the specific benefits of participating in the study
- Describe the ways in which the evaluation team will minimize the burden of participation
- Determine whether the district is sufficiently interested in the study and, if so, offer to visit the district to further discuss participation

We expect 30 of the 45 districts to express sufficient interest because of the benefits of participation. Senior staff from AIR will visit these 30 interested districts. The site visits will

allow us to provide an in-person presentation about this study, discuss the benefits and responsibilities of participation with additional district officials, and respond to any questions and concerns that they might have.

During the recruitment visit, study staff will work with the district to identify schools that meet the following criteria:

- Are either elementary schools including Grades 4 and 5, middle schools including Grades 6–8, or K–8 schools
- Do not use evaluation system components like those used in the study. That is, the school is not, on an individual school basis, implementing evaluation system components that would reduce the expected service contrast. (See criteria 2 and 3 from the section “Conducting District-Level Screening Interviews” above.) To maximize the number of schools that agree to participate, AIR will ask district leadership to convey support for the study and to set the expectation that the qualifying schools participate in the study.

Study staff will identify several incentives to participation as part of the recruitment process. One of our main inducements is that we have done a lot of the upfront investigative work to identify high quality evaluation providers in the field and have created a package for the districts to adopt. The strongest incentive we will emphasize, however, is the opportunity to take advantage of the assistance that we will provide with implementing a comprehensive educator evaluation system.

Finally, in late spring/early summer 2012, AIR will host a conference for the top 25–30 prospects to learn more about the study, discuss the content of the project in detail, and review random assignment and research-related responsibilities. One staff person per district will be invited to this conference. Attendance will not be mandatory, and districts still can participate in the study if they decline to attend the conference. Similar conferences have been used quite effectively in other random assignment studies to deepen district understanding about the study, answer questions, and solidify support for the evaluation. Such conferences also have proven to be efficient because they reduce the amount of staff time and resources needed to make visits to districts and thus reduce overall recruitment costs.

### **Negotiating Final Agreements**

Shortly after the conference, district administrators will be asked to reach a final agreement to participate, and a memorandum of understanding with interested districts will be prepared. As a condition of participation, districts also must ask eligible elementary and middle school principals to submit signed statements reflecting an intention to participate. If necessary, project staff will make additional visits to build consensus and obtain commitment from principals or other affected parties. The principal signatures are expected to be gathered shortly after the district memorandum of understanding is obtained in order to allow random assignment before the start of the 2012–13 school year.

## ***2. Procedures for Data Collection***

The district-level screening protocol is included in this package in Appendix A-2. The items on the screening protocol are mapped to the constructs that they measure in Exhibit 1. The screening protocol will be administered by project staff via telephone interviews with district personnel.

**Exhibit 1. Mapping of District-Level Screening Protocol Items to Constructs**

District-Level Screening Protocol	
Sufficient contrast between proposed system and current evaluation system	1-5c, 6-15, 17-26, 28-29, 43-44
District not currently using parts of proposed teacher and leader evaluation system	5d, 16, 27, 32-34
District context for implementing study	30-31
Data system capacity for student growth analyses	35-42

**3. Procedures to Maximize Response Rates**

Based on the previous experience of study staff, the anticipated response rate is approximately 85 percent for the districts initially screened. The following procedures will be used to ensure high response rates:

Obtaining high response rates depends in part on the quality of the instruments. The screening protocol will be pilot-tested to ensure that the questions are clear and as simple as possible for respondents to complete.

The study will offer a social incentive to respondents by stressing the importance of the data collection as part of a high-profile study that will provide much-needed information to districts and schools.

**4. Pilot-Testing Instruments**

The district-level screening protocol was pilot-tested with a small sample of respondents (fewer than 10) for two purposes—to ensure that the instrument and procedures work effectively, and to sharpen estimates of the respondent burden. Based on these considerations, the screener was pilot-tested by telephone with a convenience sample during January 2012. The individuals who participated in the pilot-testing included district-level directors of human resources or personnel services. Upon completion of the pilot-testing, the items on the screener were found to function as expected, and the time required to complete the screener questions was accurately estimated.

**5. Names of Statistical and Methodological Consultants and Data Collectors**

This project is being conducted under contract to the U.S. Department of Education by AIR. Michael Garett is the Principal Investigator, and Andrew Wayne is the Project Director. The senior task leaders from AIR contributing to the study methods and data collection are Jinok Kim, Anja Kurki, and David Manzeske. In addition, for activities associated with the classroom observations, the project includes a subcontract to Instructional Research Group (IRG). Key IRG



staff are Russell Gersten, Joe Dimino, and Mary Jo Taylor. The district-level screening protocol was developed by Anja Kurki at AIR.