Impact Evaluation of
Academic Language Intervention

■ ■ ■

OMB FORMS CLEARANCE REQUEST:

Supporting Statement Part B

DRAFT: MARCH 2017

Contents

Part B: Collection of Information Employing Statistical Methods

This package requests clearance from the Office of Management and Budget (OMB) to conduct data collection activities for a rigorous evaluation of an academic language intervention on English Learner (EL) students' and disadvantaged non-EL students' language and reading skills. The Institute of Education Sciences, within the U.S. Department of Education, awarded the contract to conduct this evaluation to MDRC and its partners Abt Associates and the Florida Center for Reading Research at Florida State University (collectively, referred to hereafter as "the study team") in September 2015.

Some research suggests that ELs and economically disadvantaged students are at particular risk for poor academic outcomes due to underdeveloped academic language skills (Kieffer, 2010). Academic language generally refers to linguistic features that are prevalent in academic discourse across school content areas that are infrequent in colloquial conversations. Specifically for this project, academic language is defined as knowledge and understanding of words and discourse found in text that forms the basis for the language of schooling. Knowledge of academic words and discourse can be taught, practiced, and demonstrated in school in oral modalities (speaking and listening) and text modalities (reading and writing).There is a growing body of work to suggest that ELs and economically disadvantaged students struggle to develop academic language proficiency that taps the content of academic texts and academic talk; the ability to think and learn like a scientist, historian, mathematician, or writer; and the skills necessary for overall academic achievement (Bailey & Heritage, 2008; Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015; Guerrero, 2004; Hakuta et al., 2000; Honig, 2010; Shanahan & Shanahan, 2008).

Although prior studies of academic language instruction provide some initial evidence of efficacious instructional practices, there is little confirmation regarding the large-scale effectiveness of academic language instruction or intervention. The goal of this evaluation is to assess the impact of an academic language intervention on EL students' and disadvantaged non-EL students' (e.g., students from low income families) language and reading skill when implemented at a larger scale. This evaluation will contribute to the knowledge base of the instructional practices that improve language and literacy outcomes.

This submission requests clearance to conduct data collection for the baseline period prior to implementing the selected academic language intervention, during the implementation year (the 2017-18 school year), and a follow-up year (spring 2019). The evaluation will examine the implementation and impact of WordGen Elementary, an academic language intervention, using a random assignment design in which participating schools in each district are randomly assigned to a treatment group whose 4th and 5th grade teachers receive training and materials to implement the treatment or to a control group whose teachers do not. The analyses for this study will draw on the following data sources: Teacher surveys, teacher and student rosters, school district records data, student assessments, and classroom observations.

1. Respondent Universe and Sampling Methods

The respondent universe for this study includes public, Title-I elementary schools within the District of Columbia or the 48 continental U.S. states serving grades 4 and 5 with student populations that include 30 percent or greater English Learners (ELs) and an additional 25 percent of non-EL students who are otherwise disadvantaged. To understand the respondent universe, the study team used the Civil Rights Data Collection (CDRC) to identify schools that meet these criteria and identified a total of 3,237 schools in 281 school districts.

We will not draw a random sample of districts or schools for the study because districts and schools must be willing to participate and implement an academic language intervention and not already be implementing a comparable academic language intervention. When selecting districts for participation,

the study team will aim to include approximately 72 schools from 9 to 12 school districts. The study sample will include all 4th and 5th grade classrooms in participating schools. Selected schools must teach ELs and non-ELs together; both districts and schools must be willing and able to support implementation of an academic language intervention and comply with study data collection needs. In the event that we identify more suitable and willing districts than we need, the final phase of the sampling process will take into account the advantage for the policy relevance of results of having a pool of sites that is as generalizable as possible by including a variety of regions, settings, and characteristics of the student body. Within each district, schools will be randomly assigned to treatment or control on a rolling basis using standard computerized procedures.

Based on preliminary analyses of the Schools and Staffing Survey (SASS) data, there are an estimated 4 classrooms (i.e., 4 teachers) per grade in grades 4 and 5, resulting in a teacher respondent universe of 576 teachers across the 72 schools in the study. Assuming a class size of 20 students per classroom, the student respondent universe is estimated to be 11,520 students. Exhibit 1 summarizes the sample size estimates for the study for each of the following proposed data collection components:

- **School District Records Data:** To determine the impact of the intervention on student outcomes, the study team will request extant data from school districts, including demographic and academic data (e.g., gender, free/reduced price lunch eligibility; EL and special education status) for students enrolled in 4th or 5th grade in each participating school in 2017-18 and 2018-19.

- **Student and Teacher Rosters:** To support the teacher surveys and student assessments and to permit tracking of students and teachers in the treatment and BAU classrooms, schools will be asked to submit rosters of the students enrolled in each classroom as well as the name of the teacher and his/her contact information.

- **Teacher surveys:** Teachers in the treatment and control group classrooms will be asked to complete online teacher surveys developed by the study team for purposes of measuring the instructional differences between language instruction in the treatment and BAU classrooms, and to measure fidelity of implementation of the intervention by teachers in the treatment group.

- **Classroom Observations:** To capture the degree to which teachers are delivering instruction that supports academic language and reading development irrespective of curriculum and assigned treatment condition, the team will collect observational data using the Classroom Assessment Scoring System-Upper Elementary version (CLASS-UE), as well as observational data to capture teachers' coverage of the intervention's curricular units and content and delivery of intervention-specific instructional strategies.

- **Student Assessments:** To estimate the impact of the ALI intervention on students' academic language skills and reading comprehension skills, the study team will administer the CALS-I test to study students in Spring 2018. The CALS-I is a measure of academic language (a constellation of the high-utility language skills that correspond to linguistic features prevalent in oral and written academic discourse across school content areas but which are infrequent in colloquial conversations) with predictive validity and high reliability. The study team will also conduct a common reading comprehension assessment, such as the Iowa Test of Basic Skills (ITBS).

Exhibit 1. Sample Size Estimates

Respondent	Data source	Number of targeted respondents	Expected response rate	Expected Number in Final Sample
Districts	School district records data	12	100%	12

Schools	Student and teacher rosters	72	100%	72
Teachers	Teacher Surveys	576	85%	490
N/A (sampling classrooms)	Classroom observations	576	37.5%	216[1]
Students	Student assessments	11,520	85%	9,792

2. Procedures for Data Collection

a. Statistical Methods for Sample selection

As described in our response in section 1, this study will include a purposive sample of 72 schools in 9-12 districts; the schools must include 4th and 5th grade classrooms in which EL and non-EL students are taught together. We will then randomly assign half of the participating schools in each district to a treatment group that receives support for implementing the intervention academic language and half to a control group that does not. We will not statistically sample the districts, schools, teachers, or students for the collection of school records data, student and teacher rosters, teacher surveys, or student assessments. Instead, we will seek to collect data for all members of the study sample.

The study team will select a sample for the classroom observations. Specifically, classroom observations will be conducted in a random subsample of approximately 216 classrooms (108 from the treatment condition, 108 from the control condition) and in two rounds, once in the fall of 2017 and once in the spring of 2018. To select the sample for classroom observations, the study team will randomly select three classrooms in each study school in the treatment and control groups, selecting two fourth grade classrooms and one fifth grade classroom in half of the schools and two fifth and one fourth grade classroom in the other half of the schools.

b. Estimation Procedures

This section describes the estimation procedures for determining the impact of the intervention on student outcomes; the service contrast (i.e., differences in instruction) between treatment and control conditions; the extent to which the intervention was implemented with fidelity. We describe each estimation procedure separately.

Impact Analyses
The study team will use two-level hierarchical models to estimate the impact of the intervention on academic language skills, reading and mathematics achievement (as measured by state tests) for EL and disadvantaged non-EL students at the end of the implementation year (SY2017-18). Analyses will also examine impacts on subgroups (by grade, gender, and prior reading level) and test whether there are impacts on achievement evident at the end of the follow-up year (SY2018-19).

Three types of data will be used to estimate impacts: outcome measures, student baseline measures, and student demographic characteristics. The outcome measures include: the CALS-I; reading comprehension, and state reading and math achievement test scores. The academic language outcome data (CALS-I) will be collected at the end of the first implementation year (Spring 2018); achievement scores will be collected at the end of the first implementation year (Spring 2018) and the follow-up year (Spring 2019).

---

[1] The study team plans to sample 3 classrooms per school.

Impact analyses of both academic language and achievement will include scores from the state reading assessments administered in the spring of 2017 as baseline covariates. Estimation models will include additional covariates such as students' EL, socioeconomic, and special education (SPED) status, attendance, race/ethnicity, gender, and age prior to the implementation of the intervention.

The estimation models will account for the randomization of schools to one of the two groups with blocking on districts, control for clustering of students within schools, and include baseline and other covariates. Each model pools across grade-levels and includes district-by-grade indicators for blocking by districts; it yields district-specific impact estimates for the intervention. The average estimate across districts or the overall impact estimate across districts and grades can be estimated by averaging these estimates. Equation 1 shows the prototypical model that will be estimated for Spring 2018 CALS-I scores and Spring 2018 and Spring 2019 reading achievement test scores; this same impact model will be estimated separately for each key subgroup of interest defined based on gender, grade-level, and prior achievement:

$$Y_{ik} = \sum_m \sum_n \gamma_{0mn} B_{mnki} + \sum_m \gamma_{1m} T_k D_{mki} + \sum_m \gamma_{2m} y_{-1ink} B_{mnki} + \sum_l \alpha_l X_{lik} + \mu_k + \varepsilon_{ik}$$

(1)

Where:

Y_{ik} = outcome measure for student i from school k;

B_{mnki} = district-by-grade fixed effect, i.e., it equals one if student i is in school k, grade n, and district m, and zero otherwise;

D_{mki} = district fixed effect, equals one if student i is in school k and district m, and zero otherwise;

T_{kn} = one if school k is assigned to the intervention, and zero if school k is assigned to BAU;

y_{-1ink} = student-level prior achievement score for student i from school k, before random assignment;

X_{lik} = student-level covariate l for student i from school k;

μ_k , ε_{ik} = a school-level and student-level random error respectively, assumed to be independently and identically distributed.

In this model, the estimated γ_{1m} represents the program impact for district *m*. The district-specific estimates can be averaged across districts, weighting each by the number of treatment group schools in that district, to yield the overall impact estimate γ_1 for the average treatment school in the sample. A two-tailed t-test will be used to assess whether γ_1 differs from zero. The model controls for student baseline test scores and student-level demographic information that will be obtained from student record data.[2] The error term structure reflects the "nested" structure of the data, which has students clustered within schools.

---

[2] It is possible that we will have different data from different districts and we will deal with this by interacting the covariate that differs across districts with district indicators in the impact analysis. One example would be different reading tests used by states in the sample. Another example would be the measure of poverty status. Some districts might provide students' free and reduced-price lunch status, while others might provide free textbook status. In both cases, the underlying construct being captured by these variables is similar but their definition/measurement is not the same.

Impact on Instruction Analyses
The study team will estimate the impact on instruction, or the differences between the treatment and BAU conditions in implementing key elements of the intervention in the implementation year (2017-18) and a year after the implementation year (2018-19) (also referred to as "service contrast").

These analyses will use data from teacher surveys administered to all study teachers in fall of 2017, spring of 2018 and spring of 2019 and classroom observations conducted in the fall of 2017 and spring of 2018 in a random subsample of treatment and BAU classrooms (see Section 1 for a description of teacher surveys and classroom observation measures).

The study team will estimate the impacts on instruction using a modified version of the impact model specified in Equation 1. Equation 2 shows the prototypical model that will be used to estimate the difference between the treatment and BAU conditions for each of a set of pre-determined service contrast measures (i.e., each measure may be an item or composites of items aggregated to a domain level derived from classroom observation and teacher survey data):

$$W_{jk} = \sum_m \sum_n \gamma_{omn} B_{mnjk} + ¿ \sum_m \gamma_{1m} D_{mjk} T_k + ¿ \sum_l \alpha_{1l} X_{ljk} + \mu_k + \varepsilon_{jk} ¿ ¿ \qquad (2)$$

Where:

W_{jk} = service contrast measure for teacher *j* from school *k,*

B_{mnjk} = district-by-grade fixed effect, i.e., indicator variable which equals one if teacher *j* is in school *k* in grade *n* and district *m* and zero otherwise,

D_{mjk} = district fixed effect, equals to one if teacher j in school k is in district m, and zero otherwise,

T_{nk} = Treatment indicator which equals one if school *k* is assigned to the intervention; zero if it is assigned to the business-as-usual;

X_{ljk} = lth teacher-level pre-treatment (baseline) measure included in the model as a covariate,

μ_k, ε_{jk} = school and teacher-level random errors respectively assumed to be independently and identically distributed.

In this model, γ_{1m} represents the service-contrast for district *m*. The district-specific estimates for a grade-level can be averaged across districts, weighting each by the number of treatment group schools in that district, to yield the overall service contrast estimate for the average treatment school in the study sample.

Implementation Analyses
The implementation analyses will examine the degree to which the intervention was implemented with fidelity. Two types of fidelity are relevant to this analysis: fidelity of implementation supports – that is, the extent to which teacher professional development and ongoing support for teachers was delivered by the intervention developer as intended – and of core intervention components —namely the extent to which classroom instruction was delivered as the developer intended by the treatment teachers. Data sources include:

- Training agenda, materials and resources used to train teachers on the intervention;
- Assessment of training attendance, site-level fidelity and classroom observation data; and
- Selected items from the teacher survey;

The study team will use descriptive statistical methods to report levels of implementation for each implementation support (e.g., number and duration of training sessions offered) and each core intervention component (e.g., instructional approach for teaching vocabulary). These levels will be compared to *a priori* developer-specified thresholds indicating dosage and delivery criteria for "high" "moderate" and "low" levels of fidelity of implementation. By comparing observed levels of fidelity to developer-defined criteria, the study team will determine the degree to which each implementation support and core intervention component was delivered with fidelity. An aggregate measure of the fidelity of implementation across the implementation supports and core intervention components, respectively, will also be reported. For example, if the developer defined high implementation fidelity as the implementation of 75 percent or more of the implementation supports with high to moderate fidelity; 80 percent of the supports were delivered with high to moderate fidelity; therefore, the intervention's supports were implemented with high fidelity.

Additional analyses will examine factors associated with implementation fidelity during the implementation year. For these analyses, the study team will extract additional extant data from the Common Core of Data (CCD) and the Civil Rights Data Collection (CRDC) database on the characteristics of the participating schools and districts. The study team will calculate correlations between teacher and school-level characteristics and the measures of fidelity calculated at the teacher- and school- levels. Factors that are associated with the ability to maintain fidelity will also be examined. Analyses will estimate logistic regression models with a binary school-level measure of high or less-than-high implementation fidelity as the dependent variable and potential associated factors as independent variables. Separate models will be estimated using both Fall 2017 and Spring 2018 data to characterize factors associated with fidelity of implementation at two different points in the implementation year. Finally, the team will extract response data from surveys and teacher attendance at training to prepare an in-depth narrative description of the major challenges that emerged.

c. Degree of Accuracy Needed

This section shows the power calculations conducted to estimate the sample size requirements of the study. These calculations are based on a desired minimum detectable effect size (MDES) of 0.15 standard deviations for the impact of the intervention on a 50 percent student subsample (EL students or non-EL students, both of which are focal subgroups for the impact analysis) at the end of the implementation year. Based on the assumptions below, we estimate that the target sample of 72 schools across 12 districts will yield a MDES of 0.145.

This section includes additional power calculations to determine the minimal detectable effect sizes (MDES) for analyses that estimate service contrast differences given the sample size based on the MDES for the impact estimates. The proposed sample yields a MDES of 0.26 for teacher survey measures and 0.30 for classroom observation measures.

Statistical Power Analysis for the Impact on Student Outcomes

Following the impact model given in Equation 1, a two-level hierarchical model with students nested within schools is used for the MDES calculation. Note that the middle-level of teachers or classrooms is

omitted in this calculation.[3] It is also important to note that although the MDES formula does not explicitly represent districts (that serve as randomization blocks and are represented by fixed effects in the impact model), the district-level variance is added to the school-level.

Specifically the following formula is used for the calculation of the MDES:

$$MDES = Mul_{J-M-1} * \sqrt{\frac{\rho_{sc}^{\iota}(1-R_{sc}^2)}{P(1-P)J} + \frac{(1-\rho_{sc}^{\iota})(1-R_{st}^2)}{P(1-P)Jn*RR}}$$

(8)

where:

Mul = multiplier based available degrees of freedom in estimation. Assuming a two-tailed test with 0.05 level of statistical significance with a two-tailed test and 80 percent statistical power;

M = number of fixed effects (i.e., district-by-grade blocks) in the sample;

J = total number of schools in the sample;

P = proportion of the participating schools assigned to a given treatment condition;

n = harmonic mean number of students per school in the sample;

RR = student level response rate, assumed to be 85%;

ρ_{sc}^{ι} = intra-class correlation (ICC) at school level;

R_{sc}^2 = explanatory power of school level covariates;

R_{st}^2 = explanatory power of student level covariates.[4]

Exhibit 4 below provides details about the parameter values used in the MDES calculations and the resulting sample size requirements—this analysis suggests 72 schools are needed to reach a MDES of 0.15 for a 50 percent student subsample (assuming half of the student students are ELs and half are non-ELs).

Exhibit 4. Parameter Assumptions for the Power Analysis and Resulting Sample Size Requirements

Assumptions for Power Analysis	
Target MDES	**0.15**
P (proportion of treatment schools in the impact estimation)	0.5
n (# of students per school, assuming 85% response rate)	
--full sample	136
--50% subsample (e.g., EL or Disadvantaged Non-EL students)	68
ρ_{sc}^{ι} **(school level ICC)**	0.173
R_{sc}^2 **(explanatory power of school level covariates)**	0.768

---

[3] Theoretical and empirical evidence has shown that one will obtain nearly identical results whether or not the middle level of the model is omitted when designing cluster-level RCTs (Zhu et al, 2012).

[4]

R^2_{st} (explanatory power of student level covariates)	0.517
Result of Power Analysis	
J (total # of schools contributing to the estimation of the impact of an intervention)	72
Total number of schools required for the full sample	72
Average number of schools per randomization block[5]	6
M (number of randomization blocks)	12
Expected MDES based on above parameters	0.145

Statistical Power Analysis for the Impacts on Instruction Service Contrast Analysis

This section presents the expected MDES for the impacts on instruction analysis given the sample size requirements determined for impact analyses. The study team plans to administer the teacher survey to all participating teachers (on average 4 teachers per grade and school), for a total of 576 teachers. Classroom observations will be conducted in approximately 216 classrooms (1.5 classrooms per school per grade level or 3 classrooms per school randomly selected for observations and balanced across both intervention and BAU conditions and grade).

To assess the statistical power for the impacts on instruction, the study team relied on the only study that reported design parameters for related teacher-level measures in Reading. Specifically, Kelcey (2013) examined teacher's reading knowledge in Grades 1 through 3 as the primary outcome measure and provided estimates of intra-class correlations (ICC) and correlations between the outcome measures and baseline measures (R^2 at the teacher and school-levels), which are used as proxies for the impacts on instruction measures under consideration for this evaluation. Exhibit 5 below presents the specific design parameter estimates obtained from this study and the corresponding minimum detectable effect size (MDES) calculations conducted separately for the survey and observations measures.[6]

Exhibit 5. Assumptions and Minimum Detectable Effect Size Estimates for Impacts on Instruction

Assumptions for Power Analysis	
J (# of schools contributing to the estimation of service contrast for a given intervention)	**72**
M (number of blocks)	12
P (proportion of treatment schools in each analysis)	0.5
n (# of classes/teachers per block, assuming 85% response rate)	
--Teacher Survey	6.8
--Class Observations	3
ρ^{ι}_{sc} **(school level ICC)**	0.12
R^2_{sc} **(explanatory power of school level covariates)**	0.07
R^2_{st} **(explanatory power of student level covariates)**	0.69

[5]Districts constitute blocks since schools will be separately randomized within each district and all relevant grades within a school will be assigned to the same condition. For this design, the study aims to have at least two schools for each of the two conditions within each block, which leads to the expectation that each block will have at least four schools.

[6]The MDES calculations also assumed a response rate of 85% for the surveys and classroom observations.

Result of Power Analysis	
MDES for teacher survey	0.26
MDES for classroom observations	0.30

Exhibit 5 suggests that the service contrast estimation will yield a MDES of about 0.26 standard deviations for teacher survey measures and of about 0.30 standard deviations for classroom observation measures. This means that, if a certain instructional practice was used by 20 percent of the teachers in the BAU condition, then the smallest true effect that this study had a good chance of detecting is about 11% (i.e., increase from 20% to 31%) if measured by teacher surveys and 12% (i.e., increase from 20% to 32%) if measured by classroom observations. Given that the service contrast measures are directly targeted by the intervention, it is not unreasonable to expect large differences between treatment and BAU conditions. For example, the Reading Professional Development evaluation (Garet, et al, 2008) found a service contrast difference for teacher-led explicit instruction with a magnitude of 0.53 standard deviations in effect size when both training and coaching were provided to second grade reading teachers.

d. Unusual Problems Requiring Specialized Sampling Procedures

No unusual problems that require specialized sampling procedures are anticipated.

e. Use of Periodic Data Collection Cycles to Reduce Burden

The data collection plan reflects sensitivity to issues of efficiency and respondent burden. In this study, the study team planned data collection at the fewest intervals possible to reduce burden but ensure the quality of the data. When possible, the team will use existing data (student records maintained by participating school districts) to limit the burden on individual students or schools. The study team will request these data from districts as few times as possible to meet study needs (fall 2017, 2018 and 2019). Student and teacher rosters will be collected just twice during the implementation year (one in summer 2017, and again in early 2018); Teacher survey data will be collected just three times (fall 2017, spring 2018 and spring 2019).

3. Methods to Maximize Response Rates and Deal with Nonresponse

Data collection procedures will include several methods to maximize response rates. These methods include:
- Identifying a school liaison within each study school to centralize coordination between the study team teachers and other school staff and allowing time to schedule "make-up" data collection opportunities to address absences (i.e., among teachers or students);
- Providing school liaisons and districts contacts responsible for sharing the rosters and school records data with clear instructions and a designated point-person on the study team with whom they will work;
- Providing clear information to teachers about the study's purpose and offering incentives (pending OMB approval) to teachers who participate in data collection activities and monitoring a toll-free number and study email address for potential participants to send questions or convey concerns.

School liaisons will play a particularly important role in maximizing response rates. School liaisons will confirm student and teacher rosters and assist the study team with the scheduling of teacher surveys, classroom observations, and the student assessment sessions. In particular, the study team will work with school liaisons to determine a target week for classroom observations in the fall of 2017 and spring of

2018, as well as a target date in the spring of 2018 for administration of the assessments. The school liaison can help the study team avoid dates when major events such as statewide assessments or test preparation, in-school professional development dates, field trips or holidays are occurring (and if needed, help reschedule in the event of unexpected absenteeism, school closings, etc.), and also help facilitate make-ups if necessary. Additional information about maximizing response rates for the roster collection, school district records, and teacher surveys is presented below.

In addition, as described in Part A, the use of technology (online survey administration, secure file transfer protocol sites) will help maximize response rates by reducing the burden of participation.

Teacher and student rosters. The study team will work with district officials and each school liaison to receive student and teacher rosters. School liaisons may submit rosters electronically using a secure file transfer portal (SFTP) or by mail/FedEx using pre-addressed shipping materials provided by the study team. The study team expects a 100% response rate in collection of teacher and student rosters because this level of data is already available in schools. Each school liaison will have direct access to a designated member of the study team to assist them with any additional questions or concerns and help to streamline the process.

School district records. The study team will work with participating districts to identify a contact person at the district level who can coordinate the submission of school district records data. The team will work with this district contact to ensure that the data specifications are clear and that the minimum data set necessary to achieve study objectives are requested. Specifically, the team will share detailed instructions about the data requests and conduct follow-up phone calls to discuss the request and ensure that the district contract's questions are answered, which will help ensure districts are able to respond to the request. Each district will have direct access to a designated data analyst on the study team to assist them with any additional questions or concerns. We expect a 100% response rate in collection of student records because this level of data will be readily available and can be collected in a cost-effective manner. Providing these data will be part of the district's commitment to participating in the study.

Teacher surveys. The study team will use multiple methods to maximize response rates. The study team will 1) use a multi-mode follow-up strategy that involves emails and phone call reminders; 2) ask school liaisons to encourage participation; and 3) offer an incentive of $25 to each teacher to complete the survey after each administration (pending OMB and district approval). Once data collection is underway, the team will have access to real-time response rates to target follow-up communications. Follow-up will consist of reminders to complete the survey using multiple modes including email and telephone calls; the team will conduct up to five follow-up reminders per respondent over the course of each field period. In addition, Information provided to teachers at the start of the survey will describe the study's purpose of the study and provide a toll-free number and e-mail address to ensure that potential respondents can easily and quickly obtain answers to questions or concerns.

If necessary, the study team will ask school liaisons to encourage teachers to complete the survey (teacher participation in the survey is voluntary). Upon submission of a completed survey, each teacher survey respondent will receive a thank you message and an electronic gift card (pending OMB and district approval) redeemable at a major retailer (e.g., Amazon).[7] Gift cards will be sent as an "email gift certificate" redeemable online. These procedures are consistent with those that have produced 85 percent or higher response rates for the surveys of teachers in similar studies – our target response rate in this

---

[7]   Incentives are contingent on OMB and district approval.

study. For example, the Impact Evaluation of Data Driven Instruction Professional Development for Teachers (OMB Control No. 1850-0924), a randomized control trial, has achieved a teacher survey response rate of 92 percent (95 percent for the treatment group; 90 percent for the control group), for an online survey of 4th and 5th grade teachers in which respondents received a $20 incentive. Similarly, in the Reading First Implementation Study (2006; OMB control number 1875-0232) response rates for K-3 teachers were 95 to 96 percent.

If the response rate of an outcome measure is below 80%, the study team will conduct a series of analyses to assess how lower than expected response rates affect the validity of the impact estimates. Specifically, the team will

- Describe the degree of non-response;
- Assess whether response rate varies by treatment condition;
- Test to see if there are systematic differences in baseline characteristics across treatment groups and controls for the respondent sample. To the degree that there are no systematic differences, one may have a high degree of confidence that non-response does not break the balance between respondents in the treatment and control groups;
- Use school records data to assess whether baseline characteristics between the respondents and the non-respondents are similar. This analysis will help assess whether the two samples differ in any systematic ways and whether the impact findings based on the respondent sample can reasonably be generalized to the full sample that includes the non-respondents.

4. Tests of Procedures and Methods To Be Undertaken

The teacher surveys were pretested with fewer than nine total respondents. The pretests assessed the clarity and content of the questions and respondent burden time. The results of the pretest were used to revise and improve the survey.

5. Consultation Outside the Agency

The following individuals have been consulted on statistical aspects of the design.

Expert	Organization	Telephone Number
David Francis	University of Houston	(713) 743-8500
C. Patrick Proctor	Boston College	(617) 552-6466
Jeannette Mancilla-Martinez	Vanderbilt University	(615) 875-9452
Julie Washington	Georgia State University	(404) 413-8340
Jeffrey Smith	University of Michigan	(734) 764-5359
David Figlio	Northwestern University	(847) 491-3395
Amy Crosson	University of Pittsburgh	(814) 865-3351
Pei Zhu	MDRC	(914) 533-4035
William Corrin	MDRC	(212) 340-8840
Leigh Parise	MDRC	(212) 340-4461

Expert	Organization	Telephone Number
Austin Nichols	Abt Associates	(301) 347-5000
Catherine Darrow	Abt Associates	(617) 520-3034
Fatih Unlu	RAND	(310) 393-0411