

COLLABORATIVE STRATEGIC READING STUDY

OMB Clearance Request—Part B

SUPPORTING MATERIALS

March 2007

PREPARED FOR:

Institute of Education Sciences
United States Department of Education
Contract No. ED-06-CO-0017

PREPARED BY:

Regional Educational Laboratory—Southwest
Edvance Research, Inc.
9901 IH-10 West, Suite 700
San Antonio, Texas 78230
(210) 558-1902
(210) 558-1075 (fax)

CONTENTS

Supporting Statement for Paperwork Reduction Act Submission: Section B.....	1
B. Description of Statistical Methods.....	1
1. Respondent Universe and Sampling Methods.....	1
Phase 1—Developing an overall pool of potential sites.....	2
Phase 2—Recruiting.....	3
2. Procedures for Data Collection.....	5
Teacher Surveys.....	6
Teacher Observations.....	7
Collaborative Strategic Reading Implementation Validity Checklist (CSRIVC)..	7
Expository Reading Comprehension Classroom Observation.....	7
Student Achievement Data Collection.....	8
Student-level Achievement and Demographic Data.....	8
School Information Sheet for Recruitment.....	9
Statistical Methodology and Stratification.....	9
Estimation Procedures/Analysis Methods.....	10
Sample Characteristics and Baseline Group Equivalence.....	10
Fidelity of Implementation.....	11
HLM Analysis for Assessing the Effects of CSR on Student Reading Achievement.....	12
Level 1 (student level).....	12
Level 2 (classroom level).....	13
Level 3 (school level).....	14
HLM Analysis for Assessing the Effects of CSR on Student Reading Achievement Within ELL Subgroups.....	15
Level 1 (student level).....	15
Level 2 (classroom level).....	15
Level 3 (school level).....	15
HLM Analysis of the Relationship between the Level of Implementation and Student Reading Achievement.....	15
Level 1 (student level).....	16
Level 2 (classroom level).....	16

CONTENTS (Continued)

Degree of Accuracy Needed.....	17
3. Procedures to Maximize Response Rates.....	19
4. Tests of Procedures to Be Undertaken.....	20
5. Individuals Consulted on Statistical Aspects of Design.....	21

LIST OF EXHIBITS

Exhibit 1. District and Elementary School Counts for 5-State REL Southwest Region	1
Exhibit 2. Phase 1 Criteria for CSR 2.1.1 Study.....	2
Exhibit 3. District & Elementary School Counts by Tier Level.....	3
Exhibit 4. District & Elementary School Counts by Tier Level.....	4
Exhibit 5. Data Collection Purposes and Responsibility.....	6
Exhibit 6. Minimal Detectable Effect Size for 80 Classrooms When Schools Are Modeled Either as Random or Fixed Effects.....	18
Exhibit 7. Minimal Detectable Effect Size for 80 Classrooms When Schools are Modeled Either as Random or Fixed Effects, Students with Previous or Current ELL Designation.....	19
Exhibit 8. Minimal Detectable Effect Size for 80 Classrooms When Schools are Modeled as Fixed Effects, and 60 Percent of Parents Return Consent Forms.....	19

SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION: SECTION B

B. DESCRIPTION OF STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

This study is an evaluation of a reading comprehension program for 5th grade students, including ELL students. The study is being conducted as a part of the Regional Educational Laboratory—Southwest contract No. ED-06-CO-0017. Thus, the region of focus includes the states served by REL Southwest: Arkansas, Louisiana, New Mexico, Oklahoma, and Texas. Exhibit 1 displays the universe of districts and elementary schools in the 5-state region. The research design calls for two sets of ten schools, each of which ideally has four 5th grade classrooms. In addition, identifying the CSR intervention’s effectiveness specifically for ELL students requires sampling of schools with high percentages of ELL students. Finally, sampling will be concentrated in urban areas to reduce travel costs associated with data collection, increase the odds of locating large elementary schools with at least four 5th grade classrooms, and increase the likelihood of identifying districts and schools with a high proportion of ELL students.

Exhibit 1. District and Elementary School Counts for 5-State REL Southwest Region

State	Districts	Elementary Schools
Arkansas	229	583
Louisiana	89	826
New Mexico	102	448
Oklahoma	596	931
Texas	1,107	4,247
Totals	2,123	7,035

Note: Table data provided by MDR 2005–2006 catalog.

The initial selection of sites for the study will be completed in two phases of recruitment. During Phase 1, a list of districts and schools in the southwest region

that meet important study criteria will be developed using data from Market Data Retrieval (MDR). During Phase 2, recruiting will begin at the sites identified. If necessary, additional sites will also be recruited until the desired sample is obtained.

Phase 1—Developing an Overall Pool of Potential Sites

During Phase 1, the Criteria in Exhibit 2 will be used to obtain a large pool of sites from which recruiting will take place. Using data obtained from MDR, a count, by state (and selected cities) of the number of districts and schools in the Southwest region that meet all of the criteria for Tiers 1–3 was developed, with the results provided in Exhibits 3 and 4 below. Based on these results, the decision has been made to recruit schools from Texas, since this is the only state in the Southwest region that provides a sufficiently large pool. There are not a sufficient number of Tier 1 schools, but an attempt will be made to obtain sample using Tier 2 schools, and then Tier 3 schools will be added to the recruiting effort as necessary. The actual list of sites obtained from MDR will include the name, address, phone number, and key contact person (such as district level Research Director, Superintendent, etc.) for each site. The site specific data obtained from MDR will be enhanced by obtaining additional information about those sites from Common Core Data (CCD) as well as information obtained from district and school websites. This list will be sent to the REL Southwest Board of Directors (which includes all 5 Chief State School Officers, principals, teachers, business leaders, etc.) and other liaisons for support in our recruiting efforts.

Exhibit 2. Phase 1 Criteria for CSR 2.1.1 Study

Criteria			
State(s)	<i>AR, LA, NM, OK, TX</i>	<i>AR, LA, NM, OK, TX</i>	<i>AR, LA, NM, OK, TX</i>
City(ies)/Demographic profile	<i>Urban</i>	<i>Urban or Large Districts</i>	<i>Urban or Large Districts</i>
Grade level	5 th grade	5 th grade	5 th grade
Number of classrooms needed at each grade level in each school	4 or more 5 th grade classes in each school	4 or more 5 th grade classes in each school	4 or more 5 th grade classes in each school
Number of students per classroom	<i>25 students per classroom</i>	<i>At least 22 students per classroom</i>	<i>At least 22 students per classroom</i>

Criteria			
Number of teachers per grade	4 teachers per school (in 5 th grade)	4 teachers per school (in 5 th grade)	4 teachers per school (in 5 th grade)
Teacher title	Must teach both English/LA and social studies	Must teach both English/LA and social studies	Must teach both English/LA and social studies
ELL/LEP enrollment requirements	<i>High ELL/LEP enrollment = 50% +</i>	<i>High ELL/LEP enrollment = 30%+</i>	<i>High ELL/LEP enrollment = 10% +</i>

Phase 2—Recruiting

Recruiting will begin by contacting the site in the initial subsample. As sites decline participation, new sites will be added to the subsample and new recruiting contacts will be made. The subsample will be modified as necessary until the required sample for the study is obtained.

Using data obtained from Market Data Retrieval (MDR), the number of schools that meet criteria for tiers 1–3 are listed in Exhibit 3. Of the 7,035 elementary schools, 26 are Tier 1 sites; 430 are Tier 2 sites; and 1,393 are Tier 3 sites. Exhibit 4 includes the same information subdivided by state.

Exhibit 3. District & Elementary School Counts by Tier Level

Tier Level	Number of Districts	Number of Schools
Tier 1	2	26
Tier 2	47	430
Tier 3	161	1,393
Total	210	1,849

Table data provided by MDR Representative, November 2006.

Exhibit 4. District & Elementary School Counts by Tier Level

Tier Level	State	Number of Districts	Number of Schools
Tier 1	Arkansas	0	0
Tier 1	Louisiana	0	0
Tier 1	New Mexico	0	0
Tier 1	Oklahoma	0	0
Tier 1	Texas	2	26
Total Tier 1		2	26
Tier 2	Arkansas	1	7
Tier 2	Louisiana	0	0
Tier 2	New Mexico	2	1
Tier 2	Oklahoma	0	0
Tier 2	Texas	44	422
Total Tier 2		47	430
Tier 3	Arkansas	2	12
Tier 3	Louisiana	0	0
Tier 3	New Mexico	7	48
Tier 3	Oklahoma	7	28
Tier 3	Texas	145	1,305
Total Tier 3		161	1,393

Table data provided by MDR November 2006

Once potential sites have been identified the REL Southwest will (in coordination with the Governing Board and other liaisons) make preliminary contacts with appropriate persons at those sites via phone. This will be followed by introductory letters, along with a brief brochure describing the study. Southwest will follow-up as necessary with sites to accurately ascertain whether the site is interested or not in participating in the study. Those sites that are interested and are a fit with the criteria will be considered “pre-qualified” sites. A list of those pre-qualified sites will be compiled and submitted to IES for review. Once sites have been identified the recruitment effort moves to the school-level. As a final step of the recruitment we will ask principals/school personnel to update information about the school by using a school information sheet (see Part A, Overview of Data Collection section).

REL Southwest will conduct telephone conversations with District Research Directors (or other comparable persons) at each of the pre-qualified sites. During those conversations REL Southwest will provide study specific details and discuss site requirement. For interested sites, meetings will be set up with schools site officials (administrators, principals, teachers, or other key staff) and informational

materials will be mailed. During the face-to-face meetings with potential sites, more detailed information about the conduct of the study will be shared, and various forms necessary for obtaining proper approval (e.g., Partnering Expectations Document, teacher consent form, informed consent forms, etc.) will be distributed. *(Forms or materials that are to be distributed to parents will be written in both English and Spanish.)*

The Partnering Expectations Document is an informal written agreement between two parties, containing the terms under which the parties will cooperate. This document will be given to districts and/or principals and will be used to inform and to gain consent for participation in the study.

2. Procedures for Data Collection

Data collection will be carried out by three organizations: REL Southwest, AIR, and RGRG. REL Southwest will have ultimate responsibility for managing data collection and ensuring quality, coordination, and timeliness; AIR staff will have primary responsibility for collecting student achievement data and administering teacher-level surveys. RGRG will be responsible for overseeing classroom observations, and REL Southwest will collect extant data and screen districts and schools participating in the study. REL Southwest is also in charge of creating teacher- and student-level rosters, updating these rosters periodically, and creating unique study IDs for all participating teachers and students.

All collected data will be processed for data entry by AIR. As a part of the general data management, AIR will track response rates, using the unique study ID numbers created by REL Southwest. AIR will be responsible for converting responses into electronic analysis files, and ultimately producing public use data sets in accordance with the requirements of the Department of Education.

Exhibit 5 shows each organization's responsibilities regarding data collection and the timing of different data collection activities. The table acknowledges the study design that includes two 1-year long implementations of the intervention. The first part of the study (including recruitment, intervention and data collection) takes place from spring 2007 through spring 2008. The second part of the study and related data collection takes place between spring 2008 and spring 2009.

Exhibit 5. Data Collection Purposes and Responsibility

Responsible Organization	Data Collection Instrument	Primary Purpose		Data Collection Schedule				
		Provide Context/Covariates	Measure Outcomes	Summer/Fall 2007	Fall 2007	Winter/Spring 2008	Fall 2008	Winter/Spring 2009
AIR	Group Reading Assessment and Diagnostic Evaluation (GRADE)		X		X	X	X	X
AIR	Teacher Survey	X			X	X	X	X
RGRG	Classroom Observation: Collaborative Strategic Reading Implementation Validity Checklist	X			X	X	X	X
RGRG	Expository Reading Comprehension Classroom Observation (ERCCO)	X				X		X
REL Southwest	Student Background Data Request	X		X		X		
REL Southwest	School Information Sheet	X		X		X		

Teacher Surveys

AIR will be responsible for administering two versions of a teacher-level survey. The fall survey focuses on teacher background and takes less than 40 minutes to complete and the spring survey focuses on the instructional context in the classroom and takes approximately 20 minutes to complete. Both versions of the survey are paper and pencil (see A1 for item by item description of the surveys).

All teachers will fill out the fall surveys as a part of CSR intervention training for treatment group teachers or an informational welcome session for control group teachers. Administering the surveys to the teachers during these scheduled meetings guarantees a high response rate. The spring teacher survey will be administered to both treatment and control group teachers on a school basis by CSR coaches and classroom observers in February-April.

Teacher Observations

Two types of classroom observations will be conducted as a part of the data collection activities: Collaborative Strategic Reading Implementation Validity Checklist (CSRIVC) and Expository Reading Comprehension Classroom Observation (used by IES).

Collaborative Strategic Reading Implementation Validity Checklist (CSRIVC)

Fidelity observations using the CSRIVC will be conducted two times per year, once prior to the Thanksgiving break and once before the end of the school year. All teachers implementing the CSR intervention (treatment group teachers) will be observed both times. These observations will be conducted by CSR coaches and other data collection personnel hired to conduct classroom observations. These observations will last approximately 30–45 minutes.

Expository Reading Comprehension Classroom Observation

All teachers participating in the study (treatment and control teachers) will be observed using *Expository Reading Comprehension Classroom Observation Protocol* (used by IES) in the spring. The purpose of this observation is to document what is happening in control classrooms (potential contamination) as well as during non-CSR instruction in the treatment condition. *Expository Reading Comprehension Classroom Observation Protocol* will be used during social studies instruction; the items relevant to comprehension instruction in the protocol require the use of expository text, which is most likely to be used during social studies. A brief list of items will be added to the protocol to document whether and to what degree group instruction took place during the observation. It should be noted that fidelity observations and instructional strategy observations cannot be conducted simultaneously. An additional 30 minutes per observation will be needed for the observer to count tallies, to complete the coversheet and observation summary items and to make sure that time segments and their initials have been entered at the top of each page of the observation protocol.

Observations will be conducted by RGRG staff members that have extensive experience with the protocol through their involvement in another IES funded study

of reading comprehension. These observers were trained by using the following methods to reach 'gold-standard' inter-rater reliability:

- Theories and descriptions of explicit instruction. Each comprehension and vocabulary item in the instrument was explained and demonstrated. A video-based format was used to show observers examples of each item and how to record tallies (i.e., an observation of a specific teaching behavior);
- Details on how to use of the observation protocol;
- Guidelines for data collection, including scheduling, observation etiquette, and how to submit the observation protocol for data entry; and
- Inter-rater reliability checks. Observers practiced coding at least three 15-minute teaching segments for each instructional domain, comprehension and vocabulary.

RGRG will also be in charge of overseeing the scheduling, administration, collection and shipping of observations to AIR for data entry. For each city one of the observers will become a lead observer who will coordinate data collection and scheduling issues. The lead observer will schedule observations for data collection staff and will make sure that coaches who also work as observers will not observe a teacher who s/he is coaching.

Student Achievement Data Collection

AIR will collect pretest and posttest student data for the study. Pretest data will be collected at the beginning of the school year, before treatment teachers have trained their students to use the CSR intervention (during the first month of the school year). Posttest data will be collected towards the end of the school year. The exact dates of the data collection will depend on the dates of CSR training and schedules of participating school districts.

The data collectors from AIR will visit each classroom participating in the study to administer the group-administered student-level tests the Group Reading Assessment and Diagnostic Evaluation (GRADE). This test will be administered both in the fall and spring (see section A1 for a description of the instrument). Administering the test is likely to require two visits to each classroom per testing. This is due to the fact that GRADE may need to be administered in two different sessions for struggling readers.

Student-level Achievement and Demographic Data

REL Southwest will request student-level achievement and demographic data from school districts as soon as possible after the recruitment has been completed. The type of the request (i.e., district-wide or a tailored school-level request) will depend on each school district's preferences. The data sets will include identifiers, because the data have to be linked to specific children in specific classrooms. Once study IDs have been created for students, the identifiers will be removed from the data set. The data will be stored in accordance with the Privacy Act of 1974.

School Information Sheet for Recruitment

REL Southwest will develop a School Information Sheet to be used for collecting school-level information for recruitment purposes. This Sheet is developed for recruitment purposes, and will be used to update/verify data collected from public sources about the schools that have shown interest to participate in the study.

Statistical Methodology and Stratification

The mission of REL-Southwest is to conduct research that is relevant for the Southwest region of the United States. This is a study about the effectiveness of CSR in 5th grade classrooms with high percentage of ELL student, targeting school districts in the Southwest. The study includes two phases, each phase including different sets of schools, teachers and students. For each phase of the study 10 schools will be recruited, each including a 5th grade with approximately 4 teachers, and each classroom including approximately 20–25 students. Our primary targets are schools with significant ELL population (30 percent or more). Recruitment of schools with high percentage of ELL students will allow a robust testing of our secondary hypothesis: whether CSR is effective in increasing ELL students reading comprehension and vocabulary skills.

The CSR study is a multi-site cluster randomized control trial, in which 5th grade students are randomly assigned to classrooms, and classrooms will be randomly assigned to the CSR condition and the comparison condition within each participating school. All 5th grade teachers/classrooms and students in the school recruited will be included in the study. Students will be excluded from data collection if their parents do not consent participation in the study. Thus the final sample will include all 5th grade students whose parents/guardians allow study

participation. No sampling is done for data collection purposes, the data collection targets all study teachers and eligible students.

Before random assignment of students is conducted, we will find out whether blocking should be used in the randomization. For instance, to guarantee a close to an equal distribution of ELL students in each classroom and between treatment conditions within each school, we may need to use ELL status as a block in randomization process.

Estimation Procedures/Analysis Methods

This study is intended to assess CSR's effects on student reading achievement through a multi-site cluster randomized control trial, in which 5th grade students are randomly assigned to classrooms, and classrooms will be randomly assigned to the CSR condition and the comparison condition within each participating school. The primary hypothesis to be tested is whether students in the CSR classrooms demonstrate better reading achievement outcomes than students in the comparison classrooms. To clarify, the primary hypothesis will be based on the full sample of ELL students and non-ELL students. Prior to testing the primary hypothesis, a series of preliminary data analyses will be conducted. In particular, there will be a careful examination of the sample characteristics and baseline equivalence of the two study groups, the level of implementation fidelity among the CSR classrooms, and the relationships between the level of implementation fidelity and student outcomes. Student data for teachers who withdraw from the study will be collected to conduct an intent-to-treat analysis.

Given the nested data structure (i.e., students nested within classrooms, classrooms nested within schools), the primary hypothesis of this study will be tested using the hierarchical linear modeling (HLM) method (Raudenbush & Bryk, 2002). Employing the HLM method, CSR's effect on a variety of reading outcome measures will be estimated by comparing students in the CSR classrooms with their counterparts in the control classrooms. In addition to full-sample analyses, tests will be conducted of the intervention's effects within the ELL and non-ELL subgroups respectively (subgroups defined by ELL status: former ELL student, current ELL student, and native speakers). Although the subgroup analyses are likely to have somewhat lower levels of statistical power than the full-sample analyses, they should still be reasonably powerful as statistical power is determined primarily by

the sample size at the highest level of aggregation; that is, the number of schools (in this case).

Sample Characteristics and Baseline Group Equivalence

The primary focus of preliminary data analysis will be on sample characteristics and group equivalence at baseline. Descriptive analyses of sample characteristics (e.g., demographic composition and attrition) will be performed with both the full sample and the two study groups separately. If relevant data are available, comparisons will also be made of the demographic characteristics of the participating schools with those of the districts or states where the schools are located, which will allow an understanding of the extent to which the sample of this study is representative of the larger population.

Although the random assignment of the study sample is expected to produce two study groups that are statistically equivalent on all measured as well as unmeasured characteristics, there may still be differences between the study groups due to random error. Moreover, post-randomization attrition of the study participants may also affect the baseline equivalence of the CSR group and the comparison group in the analytic sample. Significant baseline differences between the study groups, if not properly controlled, will lead to biased estimates of the intervention's impacts. Therefore, it is essential to examine baseline group equivalence prior to conducting the impact analyses, so that significant baseline differences can be adequately controlled through the use of covariates in the impact analyses.

Specifically, group equivalence of the analytic sample will be assessed by comparing the CSR group and the comparison group on the following student and teacher characteristics:

- *Student characteristics:* gender, race, free or reduced-price lunch status, ELL status (former ELL, current ELL, native speakers), special education status, and pretest scores
- *Teacher characteristics:* years of teaching experience, level of education, and certification

Differences in the above characteristics between the two study groups will be tested using independent-sample t-tests, and significant differences based on the t-tests will be statistically controlled in the main impact analyses.¹

Fidelity of Implementation

Implementation fidelity will be assessed through classroom observations conducted twice a year. The Collaborative Strategic Reading Intervention Validity Checklist (CSRIVC) will be used during the observations to determine the extent to which the various components of CSR are faithfully implemented in CSR classrooms. Based on the CSRIVC data, a composite measure—the Implementation Index—will be constructed to represent the overall level of implementation in CSR classrooms. During preliminary data analyses, we will explore teacher and student characteristics that are associated with the level of implementation, and assess the correlations between the Implementation Index and students' reading achievement. In subsequent analyses, the extent to which the level of implementation affects student outcomes in CSR classrooms will be assessed through more sophisticated multi-level models, as will be explicated later in this section.

HLM Analysis for Assessing the Effects of CSR on Student Reading Achievement

The primary hypothesis of this study will be tested using HLM models that compare the outcomes of students in the CSR classrooms with those of students in the control classrooms. Specifically, a three-level HLM model will be constructed with students at level 1, classrooms at level 2, and schools at level 3. In the level-1 model, student outcomes will be modeled as a function of students' pretest scores and ELL status. Although randomization will not require the use of covariate adjustments to obtain unbiased estimates of the intervention's effects, the inclusion of covariates strongly related to the outcome, particularly pretest scores, will lead to improved statistical precision of the parameter estimates (Bloom, Richburg-Hayes, & Black, 2005; Raudenbush, Martinez, & Spybrook, 2005). Moreover, the use of covariates can also adjust for significant group differences that occur by chance. In addition to pretest scores, each student's ELL subgroup membership (i.e., former

¹ The analyses will not make corrections for multiple comparisons. The purpose of these tests is to identify whether the two study groups are equivalent at baseline. Consequently it is preferable to be conservative and use t-tests uncorrected for multiple comparisons, as such corrections would make it harder to detect significant baseline differences.

ELL, current ELL, and native speakers) will also be incorporated in the student-level model, as ELL students are the target population of the intervention. The level 1 model is specified as follow:

Level 1 (student level)

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}*(Pretest)_{ijk} + \pi_{2jk}*(Former_ELL)_{ijk} + \pi_{3jk}*(Current_ELL)_{ijk} + e_{ijk}$$

where

Y_{ijk} is the outcome for student i in class j in school k ;

Pretest: the pretest score of student i in class j in school k , grand-mean centered;

Former_ELL and Current_ELL: two indicator variables representing three ELL subgroups: former ELL student, current ELL student, and native speakers), with the third subgroup being the omitted reference group; both indicators are grand-mean centered;

π_{0jk} is the average outcome of students in class j in school k ;

π_{1jk} is effect of pretest on the outcome of student i class j in school k ;

π_{2jk} and π_{3jk} are the differences in the outcome between former ELL students, current ELL students, and non-ELL students in class j in school k ; and

e_{ijk} is a random error associated with student i in class j in school k ; $e_{ijk} \sim N(0, \sigma^2)$.

The classroom average outcome estimated from the above model (i.e., level 1 intercept π_{0jk}) will be modeled as varying randomly across classrooms and as a function of the intervention at level 2, the classroom level. The level 1 slopes (π_{1jk} , π_{2jk} , and π_{3jk}) will be modeled as fixed effects at level 2, as shown in the following level 2 specification:

Level 2 (classroom level)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(CSR)_{jk} + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k}$$

$$\pi_{3jk} = \beta_{30k}$$

where

β_{00k} is the average student outcome across all classrooms in school k , adjusted for student pretest and ELL status;

CSR is an indicator variable for the intervention: $\frac{1}{2} = \text{CSR}$, and $-\frac{1}{2} = \text{comparison}$, group-mean centered;

β_{01k} is the difference in student outcome between the CSR classrooms and the comparison classrooms (i.e., intervention effect) in school k ;

β_{10k} , β_{20k} , and β_{30k} are average effects of pretest and ELL status on student outcome across all classrooms in school k ; and

r_{0jk} is a random error associated with classroom j in school k on classroom average student outcome; $r_{0jk} \sim N(0, \tau_{00k})$.

In the level 3 model, both the classroom average outcome and the CSR effect within each school (β_{00k} and β_{01k}) estimated from the classroom-level model will be modeled as random effects, assuming that both the classroom average achievement and the CSR effect differ systematically across schools. In addition the classroom average outcome and the CSR effect within each school are assumed to be potentially affected by the data collection year (indicator variable for year 1 and year 2 data collection). The effects of pretest and ELL status will be fixed at their respective grand means at the school level, as shown in the following specification:

Level 3 (school level)

$$\beta_{00k} = \gamma_{000} + \gamma_{001}\text{Year} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + \gamma_{011}\text{Year} + u_{01k}$$

$$\beta_{10k} = \gamma_{100}$$

$$\beta_{20k} = \gamma_{200}$$

$$\beta_{30k} = \gamma_{300}$$

where,

γ_{000} is the average student outcome across all schools (i.e., grand mean);

γ_{001} is the effect of a data collection year to the average student outcome across all schools;

u_{00k} is a random error associated with school k on school average student outcome;

$$u_{00k} \sim N(0, \tau_{000});$$

γ_{010} is the average CSR effect across all schools;

γ_{011} is the effect of a data collection year to the average CSR effect across all schools;

u_{01k} is a random error associated with school k on the CSR effect; $u_{01k} \sim N(0, \tau_{010})$;

and

γ_{100} , γ_{200} , and γ_{300} are average effects of pretest and ELL status on the student outcome across all schools.

Of primary interest among the level 3 coefficients is γ_{010} , which represents the intervention's main effect on the outcome across all schools. A statistically significant positive value of γ_{010} will confirm the hypothesis that students in the CSR classrooms demonstrate higher levels of reading achievement than their

counterparts in the comparison classrooms. The interpretation of the intervention's effect, however, would need to be qualified if there is a significant amount of variation of the effect across schools as indicated by a statistically significant value of τ_{010} , which would suggest that the intervention has different effects in different schools rather than having a common effect across all schools. The level 3 residuals for the intervention effect generated from the above model (u_{01k}) will further reveal in which schools CSR has a particularly strong effect and in which schools CSR has a less strong effect or no effect.

In addition to the statistical significance of CSR effect, the analysis will also gauge the magnitude of the effect with the effect size index. Specifically, the effect size will be computed as a standardized mean difference (Hedges's g) by dividing the adjusted group mean difference (γ_{010}) by the unadjusted pooled within-group standard deviation of the outcome measure.

HLM Analysis for Assessing the Effects of CSR on Student Reading Achievement Within ELL Subgroups

In addition to the full-sample analysis described above, the CSR's effects will also be tested within the following ELL subgroups separately: former ELL students, current ELL students, and native speakers, if the subgroup has enough students for the analysis. If there are too few ELL students in the sample recruited, two ELL subgroups (former ELL and current ELL) will be combined and the effects will be examined within the combined group. The specific analytic model for the subgroup analysis is similar to that for the full-sample analysis, except that the effect of the CSR intervention for ELL subgroups is included as cross-level interaction terms in the HLM model. In essence, we are interested in examining whether the CSR intervention affects the slope associated with the ELL/Non-ELL subgroups.

Level 1 (student level)

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}*(Pretest)_{ijk} + \pi_{2jk}*(Former_ELL)_{ijk} + \pi_{3jk}*(Current_ELL)_{ijk} + e_{ijk}$$

Level 2 (classroom level)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(CSR)_{jk} + \gamma_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k} + \beta_{21k}*(CSR)_{jk} + \gamma_{2jk}$$

$$\pi_{3jk} = \beta_{30k} + \beta_{31k}*(CSR)_{jk} + \gamma_{3jk}$$

Level 3 (school level)

$$\beta_{00k} = \gamma_{000} + \gamma_{001}\text{Year} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + \gamma_{011}\text{Year} + u_{01k}$$

$$\beta_{10k} = \gamma_{100}$$

$$\beta_{20k} = \gamma_{200}$$

$$\beta_{20k} = \gamma_{201}$$

$$\beta_{30k} = \gamma_{300}$$

$$\beta_{30k} = \gamma_{301}$$

The coefficients of primary interest in the subgroup analysis are γ_{010} (the main effect) as well as γ_{201} and γ_{301} , which show whether ELL students differ from native English speakers in terms of the strength of association between CSR and student outcomes.

HLM Analysis of the Relationship Between the Level of Implementation and Student Reading Achievement

In addition to comparing student achievement in the CSR classrooms and the control classrooms within the full sample and within specific ELL subgroups, the study will also assess CSR's effects by examining the relationships between the level of implementation as measured by the Implementation Index and student outcomes within CSR classrooms. Since there are only two CSR classrooms in each school, there will not be enough degrees of freedom to treat the effect of implementation as random at the school level. Therefore, schools will be modeled as fixed effects at the classroom level in a two-level HLM analysis. The level 1 model is similar to that in previous HLM analyses; however, the subscripts for each term in the model do not contain the subscript for school ("k"), because school effects are fixed. Both pretest and ELL are centered around their respective grand means in the level 1 model shown below:

Level 1 (student level)

$$Y_{ij} = \pi_{0j} + \pi_{1j}*(\text{Pretest})_{ij} + \pi_{2j}*(\text{Former_ELL})_{ij} + \pi_{3j}*(\text{Current_ELL})_{ij} + e_{ij}$$

The level 2 model incorporates the Implementation Index as the primary predictor for classroom average student outcome (level 1 intercept, π_{0j}), as well as a set of school indicator variables to control for fixed school effects (see below). It does not include the Implementation Index-by-school interaction terms because there are not enough degrees of freedom to do so with only two CSR classrooms per school.

Level 2 (classroom level)

$$\pi_{0j} = \beta_{00} + \beta_{01} * (\text{Implementation Index})_j + \sum_{g=2}^G \beta_{0g} (\text{School}_g)_j + \beta_{0(G+1)} * (\text{Teacher Characteristic})_j + r_{0j}$$
$$\pi_{1j} = \beta_{10}$$
$$\pi_{2j} = \beta_{20}$$
$$\pi_{3j} = \beta_{30}$$

where,

β_{01} is the relationship between the level of implementation and classroom average student outcome;

School_g, $g = 2, 3, \dots, G$, are (G-1) dummy indicator variables representing the G schools, with School₁ as the omitted reference school;

β_{0g} , $g = 2, 3, \dots, G$, represents the (G-1) fixed school effects for the G schools; and

$\beta_{0(G+1)}$ is the relationship between a teacher characteristic and the classroom average outcome; and

β_{10} , β_{20} , and β_{30} are average effects of pretest and ELL status on the student outcome across all schools.

The inclusion of a control variable for teacher characteristic (e.g., years of teaching experience, level of education, and teacher knowledge) will ensure that the relationships between the level of implementation and student outcomes are not confounded by the teacher characteristic. Although more than one teacher characteristic could be controlled in the above model, it is advisable that the number of teacher controls be kept to the minimum and only teacher characteristics with strong correlations with the outcome should be considered given the limited sample size.

Based on the above model, the level 2 coefficient β_{01} represents the overall relationship between the Implementation Index and the classroom average student outcome across all schools, adjusted for both student pretest and ELL status and the teacher characteristic controlled at the classroom level. A statistically significant positive value of β_{01} will suggest that the level of implementation has a significant positive relationship with student reading achievement.

Degree of Accuracy Needed

Previous research conducted regarding CSR has shown positive effects sizes between 0.20 and 0.34. Accordingly, we have designed a study that can detect a

minimal effect size of approximately 0.20 for the main treatment effect and an effect size of 0.25 for subgroup analysis including ELL students.

The power analysis assumes a design in which students are randomly assigned to classrooms, and classrooms are randomly assigned within sites, and where site (i.e., school) effects are treated as random.² Although students are randomly assigned to classrooms, the intervention is still conceptualized as taking place at classroom levels: all students in is specific classroom will either receive CSR instruction or not.

The power calculations are based on the following additional assumptions:

- *Desired Statistical power:* 80 percent;
- *Statistical significance level:* the statistical significance is 0.05 (two-tailed);
- *Number of 5th grade teachers per school:* assume an average of four 5th grade teachers per school;
- *Number of students per classroom:* assume that each classroom includes 25 students with 80% posttest response rates (i.e., that 20 students per classroom will provide both pretest and posttest data)³;
- *Proportion of teachers in treatment condition:* 50% under a balanced sample allocation;
- *School-level:* Modeled either as random or fixed effects.
- *Intra-class correlation (ICC):* it is assumed that classroom-level intra-class correlation is reduced to 0.1 as a result of student-level randomization. The school-level ICC is assumed to take a value of 0.15 in the schools as random effects model.
- *Explanatory power of the pre-test:* assume that the pre-test will correlate with the post-test at the following level: $r = 0.70$; $R^2 = 0.5$, with resultant error reduction.
- *Number of school districts:* four or fewer school districts will participate in the study.

The Exhibit 6 below includes findings from the power analyses incorporating the above assumptions. Regardless of whether schools are modeled as random or fixed effects, 80 classrooms will provide a minimal detectable effect size of 0.20 or smaller. The ability to detect minimal detectable effect sizes of this quantity is due

² Whether schools will be modeled as random or fixed effects in the statistical outcome models will depend on the final number of schools participating in the study.

³ It is reasonable to assume minimal teacher turnover, because CSR is implemented for one academic year (fall semester/spring semester) in participating schools. Most teacher turnover takes place between school years (during summer); teacher turnover that takes place during a school year is typically due to events such as pregnancy or illness. In addition, replacements for attriting treatment condition teachers will receive similar training regarding CSR as the original study teachers. Moreover, the impact of teachers discontinuing their study participation during the implementation year is minimal in terms of the power of the study, provided that both pre- and post-tests data are obtained for the students. For the current power analyses it is assumed that this will be possible.

to the additional randomization of students to classrooms, which is assumed to decrease classroom-level clustering significantly.

Due to the study's focus on ELL students, we have also calculated power assuming that on average 30 percent of the students in the classrooms either have active ELL status or have been previously identified as ELL students. Due to the fact that power in the design is largely determined by the highest level of clustering (i.e., schools), the minimal detectable effect size is still below 0.25 (assuming a baseline covariate with R-square of 0.5) even if only the ELL students (current or previously identified) are included in the student-level outcome analysis (see Exhibit 7).

Exhibit 6. Minimal Detectable Effect Size for 80 Classrooms When Schools Are Modeled Either as Random or Fixed Effects

	Schools as Random Effects	Schools as Fixed Effects
<i>R-squared</i>	<i>Two-tailed</i>	<i>Two-Tailed</i>
<i>Low (0.3)</i>	0.22	0.20
<i>High (0.5)</i>	0.18	0.17

Exhibit 7. Minimal Detectable Effect Size for 80 Classrooms When Schools Are Modeled Either as Random or Fixed Effects, Students with Previous or Current ELL Designation

	Schools as Random Effects	Schools as Fixed Effects
<i>R-squared</i>	<i>Two-tailed</i>	<i>Two-Tailed</i>
<i>Low (0.3)</i>	0.27	0.26
<i>High (0.5)</i>	0.23	0.22

Due to collection of parental consent forms, it is possible that the number of students for which parental consent is received is lower than the assumed 80 percent. Exhibit 8 shows the minimal detectable effect size, when 40 percent of the

parents refuse consent. In this worst case scenario we would have 12 students per classroom for data collection and four of the twelve students would have an ELL status. Exhibit 8 shows the estimated statistical power in this situation.

Exhibit 8. Minimal Detectable Effect Size for 80 Classrooms When Schools Are Modeled as Fixed Effects, and 60 Percent of Parents Return Consent Forms

	All Students	ELL Subgroup
<i>R-squared</i>	<i>Two-tailed</i>	<i>Two-Tailed</i>
<i>Low (0.3)</i>	0.22	0.30
<i>High (0.5)</i>	0.19	0.26

3. Procedures to Maximize Response Rates

To obtain high response rates and high quality teacher- and student-level data the following steps will be taken:

- Clear parental consent forms that explain the purpose of the study and related data collection without jargon. If possible, the school district will collect the consent forms on the behalf of REL-Southwest to increase the response rates;
- Clear explanation of study requirements to ensure that participating schools (both principals and teachers) fully understand the burden created by study participation;
- Preparation of high quality instruments that are clear and do not burden teachers excessively. The bulk of the survey items and format have been pre-tested because they come from a survey used in the Professional Development Impact study commissioned by IES that was administered to 270 elementary teachers;
- Use of monetary incentives to compensate teachers for the time used to complete surveys, \$20 for the longer Fall Teacher Survey and \$10 for the shorter Spring Teacher Survey;
- Emphasizing the prestige of participating in an important study, whose results are relevant not only for the participating teachers, but potentially for all teachers teaching English/Language Arts in classrooms with high percentages of former or current ELL students.
- Provide thorough training for the staff members who will be responsible for data collection regarding each instrument to guarantee high quality and consistency in data collection across classrooms and schools;
- Assign a staff member(s) with experience with complex data collection to be a data manager. This person will be responsible for:

- o Building and maintaining good working relationship with the school districts and school personnel;
- o Scheduling data collection;
- o Overseeing and participating in data collection in person;
- Use two sessions to administer the student-level tests to avoid testing fatigue;
- Schedule an extra day for each data collection site-visit to account for last minute unexpected changes in school/teacher schedules. The extra day will also make possible testing students who were not in school during the originally scheduled test day.

We also expect the suggested data collection procedures to help to obtain high response rates. Fall Teacher Surveys will be administered as a part of CSR training (treatment group teachers) or in an informational session regarding the study (control group teachers). The Spring Teacher Surveys will be administered in person by classroom observer or a CSR coach, as a part of a scheduled observation. The power calculations for the study assumed 20 percent attrition/non-response rate for student-level data collection to acknowledge the fact that some parents/guardians may not allow study participation and some students may refuse to participate in data collection.

4. Tests of Procedures to Be Undertaken

The Student Background Data Request is modeled on one used successfully in another study (Professional Development Impact Study) for the same purpose. Its terminology is clear to school districts so that they understand the variables we are requesting.

The items in the Fall and Spring Teacher Surveys were largely taken directly from the teacher background study used in the Professional Development Impact Study, which surveyed approximately 270 elementary teachers.

The GRADE is a widely used group-administered paper-and-pencil test. Because GRADE subtests can be administered separately, the test can be divided across two or more sessions to accommodate students' needs and class schedules. The GRADE has clear administration instructions for study staff to follow.

5. Individuals Consulted on Statistical Aspects of Design

Dr. John Hitchcock, Caliber Associates

Dr. Anja Kurki, AIR

Dr. Mengli Song, AIR

Dr. Chuck Wilkins, REL Southwest

In addition to the above, members of the TWG (listed in Section A) have provided substantial input to the study design and data collection plan.