OMB PACKAGE

2.1.2 The Effectiveness of Sheltered Instruction on English Language Learners in Georgia 4th and 5th Grade Classrooms (SIOP)

Supporting Statement Part B

TABLE OF CONTENTS
OMB Package Supporting Statement B

Supporting Statement for Request for OMB Approval of Data Collection/Needs Assessment for the REL-SE

Part B. Collections of Information Employing Statistical Methods

1. *Describe (including a numerical estimate) the potential respondent universe and any sampling or any other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, household, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection has been conducted previously, include the actual response rate achieved during the last collection.*

1.1 Potential Universe and Sampling Selection

"The Effectiveness of Sheltered Instruction on English Language Learners (ELLs) in Georgia 4th and 5th Grade Classrooms" is a study being carried out by the Regional Educational Laboratory - Southeast (REL-SE). REL-SE selected Georgia as the site for this study because it is a state with a still fairly recent (as compared to states like California and Florida), and rapidly increasing population of English Language Learners in the public school population. In 2006-2007, ELLs comprised 4.3% of all public school students and the percentage of ELLs in individual Georgia schools ranged from 0% to 73%. These figures continue to increase each year, but because the influx of ELL students is relatively recent, teaching and instructional resources for ELLs are limited. This means that there exists a strong need in Georgia for proven interventions that support ELLs and their academic achievement. Furthermore, because instructional resources are limited in Georgia, it is possible to construct a sufficiently large study sample of schools that have not had exposure to ELL teaching strategies in general, and to the SIOP model in particular.

The rationale for designing an experiment to test the efficacy of SIOP in 4th and 5th grade classrooms is motivated by an identified need – the limited support available for teaching ELLs is often focused only on the lower grades, K-3, with 4th and 5th graders receiving little to no support. Hence, a focus on 4th and 5th grade allows us to provide needed support to ELLs at grade levels that are currently underserved. In addition, whereas middle and high school students have several content-specific teachers, 4th and 5th graders still learn in the elementary setting with one teacher. Therefore, instead of having to train 'X' number of teachers who work with a given group of students, we can work with just one elementary teacher. This allows us to control for potential confounding variables that may result, for example, if a group of students were to experience a different implementation of the SIOP procedures in every content area class.

In sum, to address the question of whether or not SIOP improves the academic achievement of English Language Learner (ELL) students, selected schools must have a sufficient proportion of ELL students. In addition, they must include 4th and 5th grade teachers and students because the study design focuses on these grades.

Since writing the original study plan, a review of school demographics has led us to conclude that we need to expand our universe of potential schools in order to yield the required number of schools, teachers, and students for the study (see Section 2.2).

Therefore, to expand the original pool of schools, we first lowered the threshold from schools with ≥ 20% ELLs to schools with ≥ 15% ELLs. Next, we decided to delve more deeply *within* the districts already in the pool by including all schools within those districts that have ≥ 5% ELLs. Our decision to recruit more deeply within our existing districts rather than to reach out to new districts was motivated by current time and resource constraints – by concentrating schools within a limited number of districts, recruitment, follow-up training and data collection will be facilitated. *Therefore, the current pool of schools for the study is defined as all Georgia elementary schools containing 4th and 5th grade classes with ≥ 5% ELLs that are in districts that include at least one school with ≥ 15% ELLs.*

Based on the above criteria, the pool for the study consists of 191 elementary schools in 16 districts These schools constitute 16% of the total 1,232 elementary schools in Georgia in 2006-2007,and 38% of all the ELL students in the state in that same year.. Power analyses based on the harmonic means from this pool of 191 schools indicate that our recruitment target would need to be 88 schools to be able to detect policy-relevant effects at the student level.[1] The increase in the required number of schools (we had originally planned to recruit 64 of 95 schools) is due to the fact that as we add schools with lower percentages of ELL students to the universe of schools, the harmonic means used in power calculations decrease. In this case, the mean number of teachers per schools decreased from 10 to 7 and the mean number of ELL students per class decreased from 8 to 3. The implication is that to retain the same power to detect policy-relevant effects at the student level, the number of schools must increase. We will carefully track the numbers of schools, as well as the number of teachers, and students that we can expect to include from each school that accepts to be in the study, and monitor closely the implications for the study's power to detect effects.[2]

The 2006-2007 GaDOE data provide counts of 4th and 5th grade teachers per school, as well as ELL and non-ELL students per grade, that we used as a basis for our current estimates for the study, which will occur in 2008-2009. Table 1 below summarizes the populations for the study samples of teachers, ELL students, and non-ELL students based on the 2006-2007 data for the expanded pool of 191 schools.

Table 1
Estimates of the Population

---

[1] Due to the wide ranges in the numbers of teachers and students in schools, we use harmonic rather than arithmetic means in subsequent power calculations.

[2] An additional option, if we still find that we are not able to recruit a sufficient number of schools from the expanded universe of 191 schools within the 16 districts would be to expand once more to schools with 5% or more ELLs in districts outside of the 16. This would enlarge the potential schools to 256. We do not believe that this will be necessary at this time, however.

REL
SOUTHEAST

	5+% ELL in 16 districts
Number of schools based on 2006-2007 data	191
Enrollment range and average	242-2,312 (Average 891)
Number of 4th and 5th grade teachers	1,874
Number of ELL students in 4th and 5th grade classes	5,105
Number of non-ELL students in 4th and 5th grade classes	46,693

Strictly speaking, there would clearly be interest in estimating the effects of SIOP for a broader population of schools in the state. However, these 16 districts and 191 schools have more than one-third of the ELL students in the study. Furthermore, the clustering of schools into a manageable number of districts makes the process of recruiting and working with districts and schools feasible under the time and budget constraints of the study. The findings will not be generalizable to all schools with ELL students, however, as with many studies depending on volunteers, the findings should be a reliable representation of how SIOP performs for a substantial part of the school and teacher population.

While the target sample size (88) is less than half of the size of the population (191), we do not plan to select a sample of eligible schools because the participation rate among eligible schools may not exceed 50 percent. Therefore, we plan to make initial recruiting attempts with all of the 191 eligible schools.

We expect the recruiting effort to yield a broad variety of schools from the northern part of the state of Georgia, where most of Georgia's ELLs reside. The goal of the study, then, is to achieve internally valid estimates of SIOP's effects for a diverse group of schools with a non-trivial number of ELL students.

Recruitment of schools will begin after OMB approval is obtained for participation in the study in 2008-2009. Study team members will first contact local district superintendents to secure approval to recruit schools in their districts and to ask for support in recruitment. We will then contact principals to determine if they and their teachers are interested in participating. In these recruitment activities, we will rely on the support and endorsement from the districts and the state to encourage schools to participate. There are good reasons to expect that we can recruit an adequate sample of schools and teachers for the study within the targeted districts:

- SEA Support: The REL-SE has obtained a letter of support from the GaDOE stating their enthusiasm for this project and willingness to encourage schools and teachers to participate. GaDOE English for Speakers of Other Languages (ESOL) and Title III program personnel are very enthusiastic about the study and will provide encouragement to schools considering participation.
- Early Inquiries: As information about the study has begun to spread through informal channels around the state, the study manager housed in the GaDOE, has received inquiries from teachers and principals as well as from district and Regional Education

REL SOUTHEAST

Service Agency personnel about the possibility of participating in the study and about the SIOP model.

- ▪ Student Achievement and AYP: ELLs are the lowest-performing demographic group on most state test content areas, most grade levels. The number of schools in the state that have not met AYP solely due to the scores of the ELL group has tripled in the past year (17 in 2006, up from 5 in 2005). There is widely-accepted understanding that the pull-out instruction of the state-funded ESOL program is insufficient to raise students' academic English fluency and content area achievement levels, and support is sought for the mainstream teachers who instruct ELLs during the vast majority of the academic day.
- ▪ Professional Learning Units (PLUs) Offered: The availability of PLUs provides an incentive to participating teachers.

We will approach and solicit schools until the target number of schools is obtained. Based on the high level of interest expressed to the SERVE representative in the Georgia Department of Education, the access to and support from the Georgia ESOL Department, and the expanded pool of schools from which to recruit, we are hopeful that we will obtain the required number of schools.

The study sample based on 2006-2007 GaDOE data, and increasing the number of required schools in the study sample to 88 in order to ensure we can detect policy-relevant student level effects is presented in Table 2 below. Participants will include all 4th and 5th grade teachers who have signed agreements to participate (within schools whose principals signed agreements to participate) and all students (ELL and non-ELL) in participating teachers' classrooms.

Table 2. Sampling Plan

Universe of Available Cases		Number to be Selected (N)[a]	Selection Method
Population	*2006-2007 Counts*		
Schools	191	88 schools	50% randomly assigned to treatment and 50% to control groups
Gr 4 & 5 Teachers ▪ Treatment ▪ Control	1,874	Approx. 616 teachers ▪ 308 teachers ▪ 308 teachers	All 4th & 5th grade teachers in sampled schools who agree to participate[3]
Gr 4 & 5 ELL Students ▪ Treatment ▪ Control	5,105 ▪ 2,552 ▪ 2,552	Approx. 1,848 ▪ 924 ▪ 924	All ELL students in the sampled teachers' classes
Gr 4 & 5 Non-ELL Students ▪ Treatment ▪ Control	46,693 ▪ 23,346 ▪ 23,346	Approx. 15,400 ▪ 7,700 ▪ 7,700	All non-ELL students in the sampled teachers' classes
Gr 4 & 5 TOTAL STUDENTS ▪ Treatment ▪ Control	51,798 ▪ 25,899 ▪ 25,899	Approx, 17,248 ▪ 8,624 ▪ 8,624	All students in the sampled teachers' classes

---

[3] Participation in the study is voluntary.

REL SOUTHEAST

Once we obtain signed agreements from the principals and the teachers, we will conduct random assignment *at the school level*. While randomly assigning teachers within schools would reduce the number of schools required for the study, it would introduce the potential for cross-group contamination and would not be representative of how the intervention is usually implemented. To reduce the number of schools needed, we will use a randomized block design. We will first block by district (for any district that has four or more schools) because we do not want the effects of district policies (such as teacher hiring, compensation, curriculum, classroom texts) to be conflated with SIOP effects. Next, we will pair schools within blocks by matching on the school-level average ELL English proficiency score in the prior year. These data (Spring 2007 ACCESS scores) will be available to us from the Georgia DOE's Standards, Instruction, and Assessment (Testing) department. By matching schools on this variable, we will be using what is called an ideal stratifier, since it will reduce overall chance differences between the treatment and control groups on one of the key outcome variables we will analyze in Spring 2009 (to the extent that Spring 2007 ELL scores predict Spring 2009 ELL scores at the school level). After matching schools on prior ELL English proficiency, we will randomize one school in a pair to the treatment group and one to the control group.

1.2. Analytic Techniques

Research questions and hypotheses. There are three primary research questions for this study:[4]

1. Do SIOP-trained teachers score significantly higher on the Standards Performance Continuum (SPC) assessment than do control teachers?
2. Do ELL students in schools with SIOP-trained teachers score significantly higher than ELL students in control schools on the ACCESS assessment (Georgia's state-adopted English-language proficiency assessment)?
3. Do ELL students in schools with SIOP-trained teachers score significantly higher than ELL students in control schools on the CRCT five content areas (Georgia's end-of-year achievement tests)?

The following additional research question will also be addressed:

4. Do non-ELL students in schools with SIOP-trained teachers score significantly higher than non-ELL students in control schools on the CRCT five content areas (Georgia's end-of-year achievement tests)?

Our hypothesis in terms of teacher impacts is that in schools that implement SIOP, this training will have a significant positive impact on treatment teachers' behavior and practices, as

---

[4] Research questions that address the potential differences in SIOP impact by subgroups of ELL students are not included in this study plan because the number of schools that would be needed to support subgroup analyses of student impacts is beyond the scope of this study.

compared with schools that do not have the SIOP treatment. Our hypothesis in terms of student impacts is that the SIOP intervention will lead to positive changes in teacher practices, which will in turn, positively impact ELL student outcomes. Our hypothesis underlying non-ELL outcomes is that the approaches in SIOP are really a refinement of good teaching practice, so to the extent to which teachers implement these enhanced practices, they will also be effective with non-ELL's.

Analytic method. The impact of participating in SIOP professional development on teachers and their students will be estimated by comparing the average observed outcomes of the SIOP group with those of the control group. We will calculate these average differences using hierarchical linear modeling (Raudenbush & Bryk, 2002)[5], in order to appropriately account for the clustered nature of the data (e.g., teachers clustered within schools). This method yields a direct estimate of the average difference between the two groups on any given outcome of interest, as well as an estimate of its standard error, adjusted for clustering. If this estimate is positive and statistically significant, we can conclude that participation in SIOP professional development had a positive impact on outcomes for teachers and their students (see below for model specifications).

A central assumption of RCT designs is that randomization produces two groups (whether schools, teachers, or students) that are equivalent at the outset of a study, on average, on all measured and unmeasured characteristics potentially related to outcomes of interest, thereby allowing researchers to control for all pre-existing characteristics of teachers and students (e.g., teacher knowledge or skill) by design. Therefore, we can assume that the estimate of the impact of SIOP is unbiased with respect to pre-existing characteristics and that it is unnecessary to control for pre-existing characteristics of teachers in order to assure an accurate estimate of SIOP impact. We will, however, include covariates to improve the *precision* of our estimate, including baseline measures (e.g., school-level average ELL English proficiency) as well as background characteristics of schools, teachers, and students in our analyses, thereby improving our ability to detect significant effects.

For analyses at the student level, outcomes will be measured across two grade levels. All analyses will be done by combining the grades, in order to have a large enough sample (and therefore enough power) to detect policy-relevant effects at the student level. For student outcomes, ACCESS scores are vertically scaled across grades 4 and 5 and so pose no problems for combined analyses. Because CRCT scores are not vertically scaled, we will transform grade-level scores to standardized scores before conducting analyses. In our HLM models for each outcome, we will have random intercepts and fixed-effects dummy variables for treatment and grade. We will fit one model with a treatment-by-grade interaction, and if it is significant, we will conclude that the treatment effect differs by grade. If the interaction is not significant, we will drop the interaction term and report the combined treatment effect on 4th and 5th graders together.

---

[5] Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, CA: Sage Publications.

REL SOUTHEAST

To answer research question 4, we will run separate impact analysis on the sample of non-ELL students in treatment and control classrooms, using their CRCT scores as the outcomes and including prior school-level average CRCT scores as covariates. We will then examine and describe the patterns of estimated impacts of SIOP for each sample (ELL and non-ELL) to determine whether SIOP had positive impacts on both groups.

2. ***Describe procedures for collection of information, including: statistical methodology for stratification and sample selection; estimation procedures; degree of accuracy needed for the purpose described in the justification; unusual problems requiring specialized sampling procedures; and any use of periodic (less frequent than annual) data collection cycles to reduce burden.***

Sample selection is described above in Question #1. Estimation procedures, degree of accuracy needed, and data collection are described below.

2.1. Estimation Procedures for Teacher- and Student-Level Impacts

Teacher-level impacts. To determine whether SIOP has an impact on teachers' instructional practices (research question 1), the analysis will be a two-level hierarchical linear model, where teachers are at Level 1, and schools are at Level 2. To the extent possible, we will also examine the relationships of SIOP effects and various teacher characteristics such as: years of teaching experience and professional qualification (certification). An example HLM equation follows.

Level-1: Teacher (for teacher i in school j)

$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{graduate degree})_{ij} + \beta_{2j}(\text{certification})_{ij} + \beta_{3j}(\text{years teaching})_{ij} + r_{ij}$

Where $(\text{graduate degree})_{ij} = 1$ for teachers who have masters degrees and above; $= 0$ for teachers who have bachelors degrees;

$(\text{certification})_{ij} = 1$ for teachers who have state certifications in ESL; $= 0$ for teachers who do not have ESL certifications;

$(\text{years teaching})_{ij}$ is a continuous variable of the number of years teachers have been teaching in Georgia;

and r_{ij} is the random effect at level-1 assumed to be normally distributed, with a mean of 0 and variance of σ^2.

Level-2: School

$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SIOP})_j + \gamma_{02}(\text{AYP})_j + u_{0j}$,

where $(\text{SIOP})_j = 1$ for schools assigned to the treatment group; $= 0$ for schools assigned to the control group;

$(\text{AYP})_j$ is a continuous variable of the school's annual yearly progress in the prior year as a pre-test covariate to improve statistical precision;

and u_{0j} is the random effect at level-2 assumed to be normally distributed, with a mean of 0 and variance of τ.

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

$\beta_{3j} = \gamma_{30}$

In this equation, the main effect of SIOP on the teacher outcome Y_{ij} is represented by γ_{01}, taking into account standard demographic and performance characteristics of schools and teachers.

Student-level impacts. For estimating effects on ELL students (research questions 2 and 3), we will be modeling a three-level HLM, where students are at Level-1, teachers are at Level-2, and schools are at Level-3. To increase precision, we will include student characteristics in analyses.[6] An example HLM equation is as follows:

Level-1: Student (for student i in classroom j in school k)
$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{Parent Education})_{ijk} + \pi_{2jk}(\text{Female})_{ijk} + \pi_{3jk}(\text{Black})_{ijk} +$

$\pi_{4jk}(\text{Hispanic})_{ijk} + \pi_{5jk}(\text{Other Race})_{ijk} + r_{ijk}$

Level-2: Classroom
$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{graduate degree})_{jk} + \beta_{02k}(\text{certification})_{jk} + \beta_{03k}(\text{years teaching})_{jk} + u_{00jk}$

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

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

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

$\pi_{4jk} = \beta_{40k}$

$\pi_{5jk} = \beta_{50k}$

Level-3: Schools
$\beta_{00k} = \gamma_{000} + \gamma_{001}(\text{SIOP})_k + \gamma_{002}(\text{AYP})_k + \gamma_{003}(\text{PRETEST})_k + u_{0k}$

$\beta_{01k} = \gamma_{010}$

$\beta_{02k} = \gamma_{020}$

$\beta_{03k} = \gamma_{030}$

In this equation, the main effect of SIOP on the student outcome Y_{ijk} is represented by γ_{001}, taking into account standard demographic and performance characteristics of schools, teachers, and students.

Results from the impact models will be presented in tables that include the regression-adjusted mean value of the outcome among the treatment group, the regression-adjusted mean value of the outcome among the control group, the estimated impact of SIOP on the outcome, the standard error of the estimated impact, the effect size of the impact, and the statistical significance of the impact.

---

[6] For research question 4, a separate impact model will be run for the non-ELL student sample.

REL
SOUTHEAST

2.2. Degree of Accuracy Needed

For the power analyses described below, an effort was made to balance the competing needs to stay within the available study resources and to provide sufficient power to detect small effects at the student level, since we are considering student impacts after only one year of implementation of teacher instructional practices. As discussed below, we considered a minimum detectable effect size (MDES) at the teacher level to be .40 and at the student level to be .20.

For teacher-level impacts. How many teachers and schools are needed to adequately detect an impact of SIOP treatment on teacher practices and behaviors? To answer this question, we ran a two-level cluster randomized trial (CRT) power analysis using the W.T. Grant Foundation's Optimal Design (OD) software (v. 1.77) (http://sitemaker.umich.edu/group-based/home).

From the 2006-2007 Georgia database of schools, we calculated a harmonic mean of the number of 4th and 5th grade teachers present at each school. The harmonic mean is 7 teachers, with a range of 1 to 31 teachers. This harmonic mean was used with the OD software to calculate how many schools are needed to determine teacher impact. The use of the harmonic mean takes into account the variability in the number of teachers within each school.

Figure 1 shows sample size calculations for an unconditional model ($R^2 = 0$)) with a two-tailed test at the .05 significance level. Although we will have some teacher-level background characteristics that we will use as covariates, we have run the most conservative scenario here by setting R^2 to 0. Under the assumption of an intra-class correlation of .15 (i.e., 15% of the variance in teacher practices is between schools and the remaining 85% among teachers within any given school), we will need at least 56 schools (28 SIOP/ 28 control) to achieve 80% power to detect an effect size of .40. Our hypothesis is that teachers need to change their practice by this magnitude (MDES of .40) in order for SIOP to have an impact on students. Therefore, our sample size of 88 schools (based on power calculations at the student level below) is clearly sufficient to detect the effects we are proposing at the teacher level.

REL
SOUTHEAST

Figure 1
Teacher Impacts and Power

$\alpha = 0.050$
n = 7
$\delta = 0.40, \rho = 0.15$

Power

Number of schools
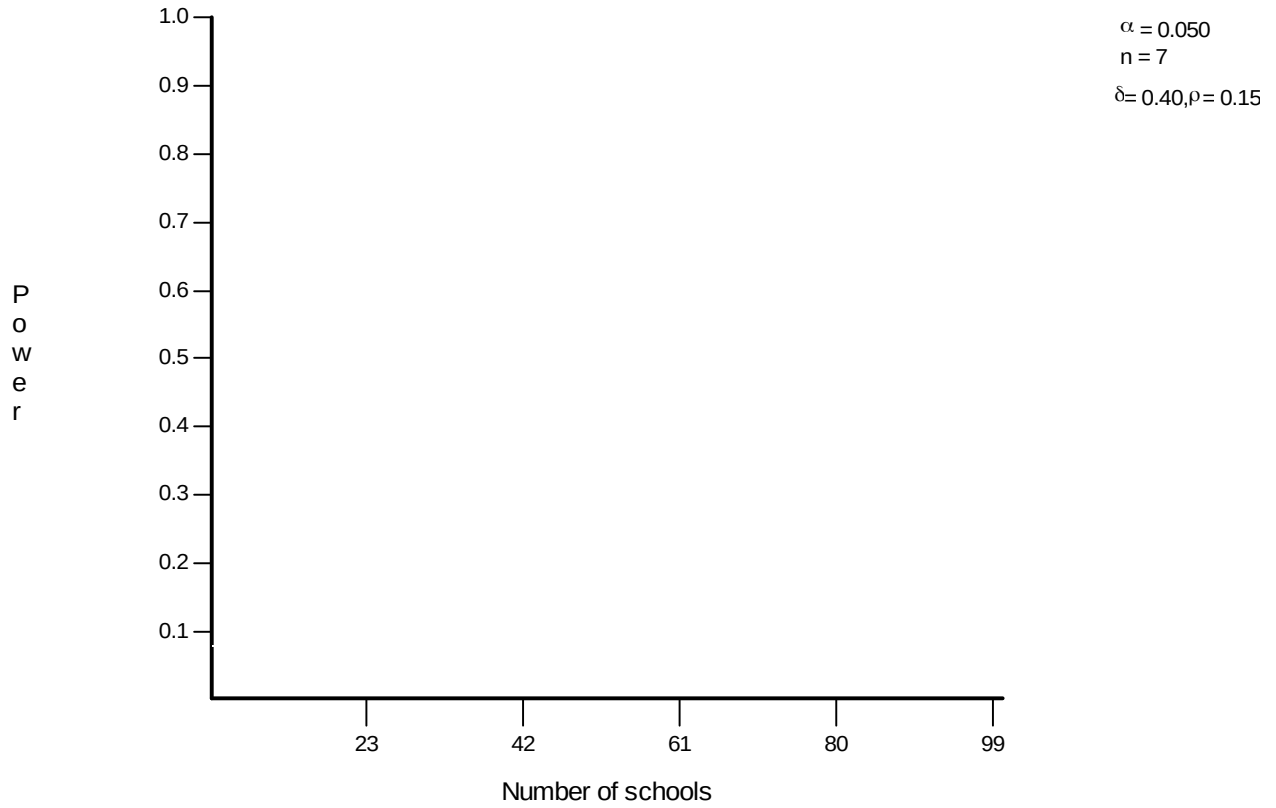
For student-level impacts. To obtain a minimum detectable effect size of 0.20 in student-level impacts, our estimates suggest that a sample of 88 schools is required.[7] Our power calculations were based on the following assumptions:
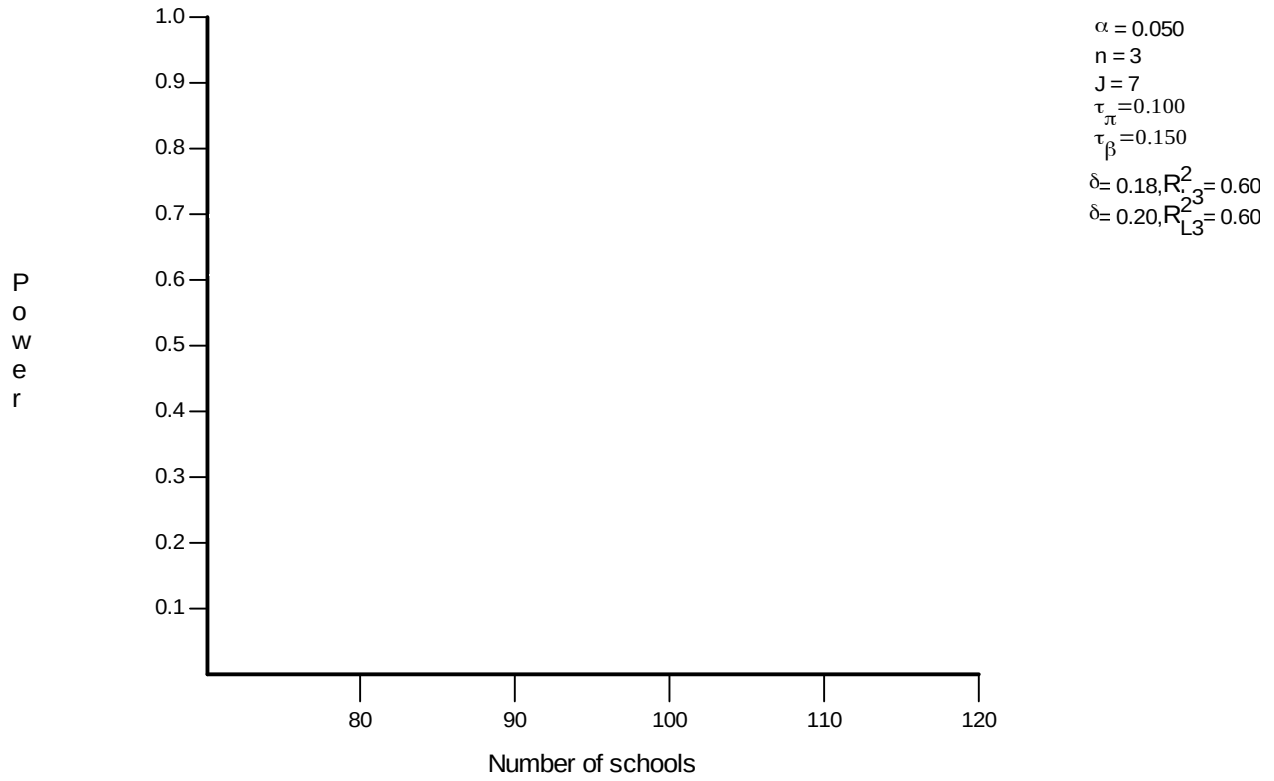
- **Participating schools will have an average of three ELL students per classroom and seven classrooms per school.** These assumptions are based on direct estimates from the population of eligible schools using data from the Georgia database of schools, which indicate that the harmonic mean is three ELL students per classroom, ranging from 1 to 20 students, and seven classrooms per school, ranging from 1 to 31 teachers.
- **The intraclass correlation across schools will be .15, and the intraclass correlation across classrooms will be .10.** Research on ELL students suggests great variability of student outcomes, where individual characteristics matter more to student outcomes than classrooms or schools. Thus, we hypothesized that the variation across classrooms and schools would be smaller than at the individual level. We also based our assumptions for classroom- and school-level ICCs on the work of Schochet (2005) and Raudenbush and Liu (2001), and on recommendations from Schochet (personal communication, 2007).
- **The correlation between the pre-test and post-test will be 0.77 (R^2=.60).** This is a reasonable assumption given the use of state test scores in the study. Prior research (Bloom, Richurg-Hayes, and Black, 2005) and recommendations from Schochet (personal communication, 2007) confirm that this estimate is within the typical range.
- **Eighty percent power and a two-tailed test of significance at the .05 level**

We explored the implications for sample size if we were to reduce the MDES to a lower level. While an MDES of .20 necessitates 88 schools, an MDES of .18 with 80 percent power requires 108 schools. However, given the already large sample size that we will need to recruit, lowering the MDES does not seem feasible at this time. We would argue that an MDES of .20 is already fairly small, and the costs of lowering it would be substantial. Based on all of the above considerations, we propose to keep the MDES of the study at .20 for student outcomes, implying a sample size of 88 schools.

---

[7] For research question 4, a separate impact models will be run for non-ELL students. Adding a substantial number of students at level 1 increases our ability to detect smaller effects – power calculations indicate that with an estimated 25 non-ELL students per teacher (as opposed to three ELL students), and under the same assumptions as the power calculations done for the ELL student sample, we will have 91% power to detect an MDES of .20.

REL SOUTHEAST

Figure 2
Student Impacts and Power

$\alpha = 0.050$
$n = 3$
$J = 7$
$\tau_\pi = 0.100$
$\tau_\beta = 0.150$
$\delta = 0.18, R^2_{.3} = 0.60$
$\delta = 0.20, R^2_{L3} = 0.60$

Power

80   90   100   110   120

Number of schools

REL
SOUTHEAST

2.3. Data Collection

Data collection will cover the full universe of 88 schools in the sample. Primary data will be collected from all 616 treatment and control teachers in the form of observations and survey data. Classroom observations, using the Standards Performance Continuum (SPC), will take place in Spring 2009 and will be conducted by SERVE data collectors. We will conduct a web-based survey of all of the participating 4th and 5th grade teachers. This survey will yield information on background data not available from the GaDOE database, as well as data on the depth and breadth of teachers' ESOL professional development (above and beyond SIOP), the amount of support for teaching ELLs in their school, and the types of materials they use in teaching ELLs. The emphasis of these questions is on activities that support high-quality instruction of English Language Learners. Since the treatment involves a substantial level of professional support, it will be important to document the level of support available to teachers both in treatment schools and in the absence of treatment, i.e., in control schools. Data collected on the teacher survey will be used for two distinct purposes: a) to construct covariates for use in the impact analyses, and b) to describe the counterfactual. The survey will be piloted prior to being used in the study in Spring 2009.

Data for students in the 88 schools will be collected from extant student records located at the Georgia DOE. These data will include:

1. Assessing Comprehension and Communication in English State to State for English Language Learners (ACCESS for ELLs), which measures four language component skills: Listening, Speaking, Reading, and Writing – We will obtain ELL students' scores on ACCESS (reading, writing, listening, and speaking) – approximately 1,848 students; and
2. Criterion Referenced Competency Test (CRCT), Georgia's state subject area achievement tests – We will obtain CRCT scores of ELL and non-ELL students (in reading English/language arts, mathematics, science, social studies) – approximately 15,400 students.

We will also collect a wide range of data from the extant data sources pertaining to school, teacher, and student characteristics, such as:

1. School characteristics, such as enrollment size, total number of teachers, number of teachers with Georgia ESOL certification, percent of students receiving free/reduced school lunch (from the Georgia Department of Education database);
2. Teacher characteristics, such as professional certification, years of teaching experience, and other professional development activities received during 2007-2008 (from the Georgia Department of Education Certified/Classified Personnel Information (CPI) database and the local school district databases, and a teacher survey); and
3. Student characteristics, such as age, gender, grade (4th or 5th grade), LEP/ ELL status, ethnicity, and first language (from the Georgia Department of Education database).

REL SOUTHEAST

The data collection plan reflects sensitivity to issues of efficiency, accuracy, and respondent burden. Wherever possible, information will be gathered from existing data sources rather than imposing additional burden by collecting primary data. A key consideration in the collection of student achievement data via existing administrative records as opposed to administering standardized tests to participants was to minimize evaluation costs and reduce respondent burden. Similarly, we restricted teacher survey content to areas not covered by the existing state administrative records in order to minimize respondent burden and are planning to utilize a web-based format in order to reduce paperwork and facilitate completion and submission.

3. **Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.**

Recruitment of Participants

We will hold a meeting for high-level state administrators to inform them of the study design and plans for conducting recruitment. At that time, we will ask for their support and will also elicit strategies for recruitment and retention of participants. Following this meeting, we will contact district school superintendents by mail (informational packet) and telephone (follow up, shortly after sending packet). We will elicit information from superintendents about the most effective way of reaching schools within their district (for example, is a meeting with all elementary school principals preferable to several individual meetings at schools in that district?). Next, we will contact principals by e-mail and telephone to provide more information about the study and opportunities for them to ask questions about the study. Once the principals have signed agreement forms, we will send teachers an informational e-mail and/or regular mail packet (depending on the district's technology use) with contact information so that they may learn more about the study and ask questions of researchers. If necessary, SERVE will send a recruitment team to conduct an informational meeting. With or without such a meeting, teachers will meet within their schools to discuss the study and to sign agreement forms. Non-respondent teachers or schools will receive a reminder e-mail and then phone calls in order to obtain our desired sample of 616 teachers and 88 schools.

Collection of Survey and Observation Data

In general, we will use the following methods to minimize attrition and maximize response: a) establish positive relationships with respondents and school staff; b) send advance letters/emails; c) establish efficient and flexible timeframes for scheduling observations and responding to surveys; d) contain burden by not asking respondents for information that can be obtained reliably elsewhere; and e) make multiple attempts at follow-up in cases of nonresponse.

In Spring 2009, prior to the observation visits, we will send an update letter (or email), bringing participants up to date on the progress of the study (this letter/email will also serve as a reminder to the control group teachers participating in the study). At the beginning of the study,

REL SOUTHEAST

we will ask for one staff member at the school to be the contact person for the study. This person will assist us in ensuring that teachers fill out the survey and have it ready when the data collector comes to conduct the classroom observation. Classroom observations will be scheduled by phone and e-mail to occur no sooner than 30 days later, depending on available technology. If teachers have not filled in the survey by the time the data collector arrives at the school, the data collector will provide a paper copy, along with an envelope, and will obtain the email address of the teacher(s) so that the study team can follow-up with those teachers soon afterward. The school's study contact person will also be apprised of those teachers who have not responded and will be asked to assist in obtaining responses.

Although we expect these efforts to be highly successful, with a sample of this size, some degree of non-response is very likely. To the extent that the level of non-response is substantial, we will test to see if non-respondents are significantly different from respondents based on their baseline characteristics, and if necessary we will re-weight the analysis sample to account for these differences. Similarly, we will examine whether there is significant differential attrition between the treatment and control groups, and again re-weight the sample if such differences are found.

4. ***Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of data.***

Student Data

Student data are to be taken from state student records: ACCESS for ELLs (reading, writing, listening, and speaking) and Criterion Referenced Competency Test (CRCT).

Teacher Data

Standards Performance Continuum (SPC). The SPC observation protocol was designed by a research team at the University of California-Berkeley's CREDE center (Doherty, Hilberg, Epaloose, & Tharp, 2002). It has been used in at least eight published studies and is supported by CREDE staff and information online.

Teacher Survey. Teacher survey to be administered to teachers in the Spring 2009 to collect background data not available from the GaDOE database, as well as data on the depth and breadth of their ESOL professional development (above and beyond SIOP), the amount of support for teaching ELLs in their school, and the types of materials they use in teaching ELLs. Some items were adapted from questions taken from the Reading First Implementation Study survey (Grade 2-3 Teacher Survey, U.S. Department of Education, Policy and Program Studies Service), which was reviewed by the Reading First Implementation Study's Technical Work Group. Items were slightly modified to reflect context and/or name of intervention, or to reflect appropriate time span and/or frequency. The entire survey for this study will be reviewed by the

technical work group for this study. The survey will be piloted prior to being used in the study in Spring 2009.

5. ***Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.***

REL-SE Project Director: Ludwig D. van Broekhuizen	336-315-7402
Co-PI: Carolyn Layzer, Abt Associates Inc.	617-520-3597
Co-PI: Micheline Chalhoub-Deville, UNC-Greensboro	336-334-3472
Study Manager: Kimberly Anderson, SERVE Center	404-657-6174
Task 2 Methodological Leader: Stephen Bell, Abt Associates Inc.	301-634-1700