Regional Education Laboratory (REL) Pacific: Random Assignment of Evaluation of Principles-Based Professional Development to Improve Reading Comprehension for English Language Learners (Task 2)

OMB Supporting Statement, Part B

Contract No: ED 06 CO 0024

Prepared for

U.S. Department of Education
Office of the Deputy Secretary
Policy and Program Studies Service
400 Maryland Avenue SW
Washington, DC 20202

Prepared by

Pacific Resources for Education and Learning
900 Fort Street Mall, Suite 1300
Honolulu, Hawai'i 96813

Berkeley Policy Associates
440 Grand Ave, Suite 500
Oakland, CA 94610

June 20, 2007
(Revised December 5, 2007)

Table of Contents

Supporting Statement for the Paper Reduction Act

APPENDICES FOR PART B

Appendix B: Teacher Survey
Appendix C: Impact Survey
Appendix D: Principal Survey
Appendix E: Teacher Focus Group Discussion Guide
Appendix F: Teacher Interview Discussion Guide
Appendix G: PREL Staff Focus Group Guide
Appendix H: Professional Development Observation Protocol
Appendix I: SIOP Classroom Observation Protocol
Appendix J: Consent Forms/Information Sheet
Appendix K: IRB Approval Letter
Appendix L: Letters of Interest
Appendix M: Sample MOU
Appendix N: Data Security Policy and Procedures

Note: Appendix A is referred to in Part A of the Supporting Statement, and is not referred in this Part B of the Supporting Statement

List of Exhibits

Introduction

This submission is a request for approval of data collection instruments that will be used to support the Evaluation of the Random Assignment of Principles-Based Professional Development to Improve Reading Comprehension for English Language Learners (Pacific-CHILD). The Pacific Communities with High-performance in Literacy Development (Pacific CHILD), is a professional development program for 4th and 5th grade teachers of secondary English Language Learners (ELL) in the Pacific region, developed by Pacific Resources for Education and Learning (PREL) with funding from Regional Education Laboratory-Pacific (REL-P). The Pacific CHILD program of professional development uses research-based instructional strategies appropriate for schools across the Pacific region. The Pacific CHILD program focuses on:

1. Using *informational* text to build reading comprehension skills.
2. Building the capacity of all students to use three reading comprehension strategies (vocabulary acquisition; question generation; and text structure) to improve reading achievement.
3. Improving pedagogy with targeted classroom organization and management practices (differentiated instruction).
4. Creating a format of instruction for 100% student engagement across the continua of reading skills and English language proficiency (interactive learning).
5. Refining practice in and with existing reading/language arts curriculum and texts.
6. Standards-based instruction, with an emphasis on closing the achievement gap between language minority and language majority students.

This study consists of two primary objectives. The first objective is to a) to determine the impact of the Pacific CHILD professional development for teachers in terms of their content knowledge, self-efficacy, and pedagogical skills, and b) to measure the impact of the Pacific CHILD professional development for teachers on student reading achievement. The second primary objective is to examine the extent to which schools and teachers receive the Pacific CHILD training and support as intended and the extent to which the Pacific CHILD model is implemented as intended. This study will also serve to inform future program improvement and replication of the Pacific CHILD program. Pacific Resources for Education and Learning (PREL) and its subcontractor, Berkeley Policy Associates (BPA), are conducting this study for the Institute of Education Sciences (IES) of the US Department of Education. While PREL delivers Pacific CHILD professional development, BPA serves as an independent evaluator. All data collection activities related to evaluation will be carried out by BPA staff.

The study adopts a cluster random assignment research design in which the unit of random assignment is the school. Approximately 50 elementary schools in Hawai'i, the Commonwealth of the Northern Mariana Islands (CNMI), and American Samoa will be recruited for this study. The 50 schools selected and their participating 4th and 5th grade teachers will be randomly assigned to treatment and control conditions. Participating teachers in the 25 treatment schools will begin the two-year Pacific CHILD professional development program during the 2007-2008 school year in American Samoa and CNMI and in the 2008-2009 school year in Hawai'i. Professional development services will be available to control schools and teachers after a 2-year embargo period in an effort to make participation in the study less of an obvious burden for control schools, which otherwise might feel that they do not benefit from the study. Data collection activities to support the study will begin only upon receipt of OMB approval.

Supporting Statement for the Paperwork Reduction Act

PART B
Collection of Information Employing Statistical Methods

1. Universe of Respondents/Sampling Plan

Universe of Respondents

The universe of respondents includes all 4th and 5th grade teachers and students in the public elementary schools in three jurisdictions (the State of Hawai'i, the Commonwealth of the Northern Mariana Islands (CNMI) and the Territory of American Samoa) in the Pacific that are participating in the evaluation. Table 1 shows the estimated numbers of schools, teachers, and students within three regions in the Pacific. The schools deemed eligible for participating in study will meet certain minimum criteria discussed below. The target population of the study is defined as all 4th and 5th grade teachers and their students in the eligible schools. The study is designed to provide findings that are generalizable to the target population, and also aim to provide present broad policy implications relevant to all elementary schools and 4th and 5th grade teachers in these three jurisdictions.

Table 1
Estimated Numbers of Schools in Target Jurisdiction

	Hawai'i	American Samoa	CNMI
Total Population	1,221,885	57,300	69,200
Number of Public Schools	285	29	20
Number of Public Schools with Grades 4 and 5	198	23	12
Number of Teachers in 4th & 5th Grade Teachers	1,024	99	85
Number of 4th and 5th Grade Students	25,543	2,133	1726

Source: School, teacher, and student numbers are calculated by BPA staff based on data collected from the Departments of Education in Hawai'i, American Samoa, and CNMI.

Sampling Plan

The study adopts a cluster random assignment research design, in which the unit of random assignment is the school. The sample of approximately 50 schools is selected from three target jurisdictions across the Pacific: the State of Hawai'i, the Commonwealth of the Northern Mariana Islands (CNMI) and the Territory of American Samoa, with 25 of the schools in treatment and 25 in control condition. It should be noted that a probability sample of schools across the Pacific is not feasible for this setting due to the logistical constraints posed by the geographic dispersion of schools throughout the Pacific Region. The only way to cost-effectively evaluate the intervention being studied is by clustering schools in geographic entities and on specific islands.

Among the schools that met minimum criteria discussed above, we will select schools across the three target jurisdictions for recruitment based on a set of additional criteria, which will include:

- Schools with the sufficient numbers of 4th and 5th grade teachers (at least two teacher per school)
- Schools with a sufficiently large percentage of English Language Learners (at least 10 percent in Hawai'i, not an issue in the other entities)
- Schools that are not already delivering intensive professional development in reading comprehension through another intervention that is not compatible with Pacific CHILD (e.g. Direct Instruction or Corrective Reading).
- Schools with SEA- and district-level support for participating in the study

All of these schools will be approached and recruited. If more than the required number of schools volunteers to participate in the study, we will increase the sample by approximately 5-6 schools as an extra safeguard against future attrition. Beyond that figure we will randomly select schools to be excluded from the study from those that volunteer to participate.

All 4th and 5th grade teachers in each of the program schools will be recruited into the principles-based professional development model for two years. Control group teachers will not receive these services for 2 years, after which they will be eligible for them as well. By promising control schools that they will gain access to similar professional development services after a two-year delay, we expect to increase the willingness of these schools to participate in the study and to minimize the degree to which control school officials seek out other professional development services on their own.

Table 2 reflects how the sample of schools, teachers, and students may be distributed across the Pacific region. The number of schools in each jurisdiction was determined first by assigning Hawai'i's 24 of the 50 schools, and then by including a minimum of two schools per jurisdiction, up to the proportion of schools in each jurisdiction relative to the total number of Pacific region schools outside of Hawai'i.

Table 2
Sample Distribution Across the Target Jurisdictions in the Pacific Region

	Schools	Teachers	Students
Hawai'i	24	124	3,096
American Samoa	15	53	1,644
CNMI	10	79	1,657
Total	50	256	6,397

*Based on the data provided SEAs and recruitment information available at the time of writing.

Anticipated Response (Participation) Rate

Explanations for anticipated response rates are requested for this section. Because all of the schools that meet the criteria discussed above will be approached and solicited until the target number is obtained, the question on the rate of response (i.e. the rate of participation in the study) is not quite relevant. Still, there may be a concern that the larger the number of schools that are not responsive to the recruitment, the more schools may need to be approached in order to reach the target number, and the less likely they will be representative of the Pacific Region. While this is an important theoretical consideration, it is possible to look at this from another perspective, namely that in which the school is considered a potential consumer of professional development services. The relevant sample for a study of this kind consists of the schools in the region that *would consider acquiring the P-CHILD professional development model if it were offered to them.* Thus, the only source of potential bias is schools that would do so in the absence of

the study, but would not because of the random assignment (and the likelihood of ending up in a control group). Based on our conversations with schools, this is not a major issue for them, especially since control schools will receive a delayed treatment. When schools refuse to participate in the study it is mostly because they are too busy with other school priorities, they cannot get their teachers to buy into it, they cannot pay for their teachers' release time, or they are already engaged in another professional development intervention. These schools would not have participated in Pacific CHILD anyway, so it is not necessary (or helpful) for the study results to be generalizable to them. In terms of the numbers of schools, at the time of writing this, we have successfully obtained the agreement to participate from 10 of 12 eligible schools from CNMI and 15 of 23 eligible schools in American Samoa. We plan to recruit Hawai'i schools (with the target range of 24-30 schools, with the minimum of 24 as a goal). Our discussions with the Area Superintendents in Hawai'i suggest that we would be recruiting from 60-70 eligible schools.

In terms of participation rates among teachers, we expect an 80 percent or higher rate of participation amongst teachers recruited at schools that agree to participate. In the case of Hawai'i, high teacher participation is expected because the school's decision to participate involves prior consultation with teachers. We do not expect, particularly for Hawai'i, that a principal would unilaterally decide to participate, but the participation decision itself reflects teachers' strong interest in the program. In other jurisdictions, the decision to participate in the Pacific CHILD program can be made primarily at the school level partly due to the top-down nature of the education system in those jurisdictions. (Note, however, that teachers will have freedom to decline to participate in the evaluation study components and informed consent process will ensure that their participation in the data collection is voluntary.)

In terms of response rates for data collection activities (from teachers and principals), we aim at and expect an 80 percent response rate. This estimate is based on several factors that will be discussed in Section 3 below.

Justification for Sample Size

In making decisions about sample sizes, our goal is to attain minimum detectable program impacts that are sufficiently small to allow reliable detection of program effects that are large enough to be meaningful from a policy perspective. If we increase the number of schools in the sample, the minimum detectable effect size (MDES) would be reduced, thereby increasing the statistical power of the study; therefore, we usually try to maximize the sample size. However, at the same time, the benefits of increased statistical power need to be assessed in terms of their usefulness for policy making. For example, the ability to detect a small, yet statistically significant impact (say, a 1% increase in a teacher content knowledge measure), may be inconsequential if such a small improvement is regarded as having no meaningful effect in policy discussions. Thus, the research design must be able to detect the smallest impact that is considered large enough from a policy perspective. The threshold for choosing an appropriate level of statistical power and sample size for the study depends on multiple factors, including, among other things, the cost of implementing a specific intervention (including the opportunity cost of alternatives), concerns about face validity, the cost of the study itself, and the need to meet rigorous research standards.

For this school-level cluster random assignment study, we propose the target sample size of 50, which is expected to yield of the minimum detectable effect size (MDES) of 0.19 for student outcomes and 0.39 for teacher outcomes (assuming that regression estimation with baseline covariates will reduce the effects of clustering by lowering the intraclass correlation to 0.05).   (The discussion on our statistical power analysis and MDES is presented in more detail in the next section). We argue that MDES of 0.19 for student outcomes and 0.39 for teacher outcomes are small enough for the full sample. Impacts smaller than this MDES level would likely be considered meaningless from a policy perspective. For subgroup analyses based on the subsample of 25 schools, MDES is estimated to be larger (0.28 for student outcomes and 0.57 for teacher outcomes), but we argue that they are generally small enough to draw reliable, policy-relevant conclusions.

While the study does not provide as much power at the teacher level, we do not view this as a serious concern. In fact, we do not expect that impacts on teachers smaller than this MDES level would be meaningful from a policy perspective, since it is unlikely that such a small impact on teacher knowledge or classroom practice would produce a meaningful improvement in student achievement, which ultimately is the only outcome that determines whether or not the program is worth continuing and disseminating across the region. Proximal teacher and classroom-level outcomes are important for understanding the program impacts and refining the program, but detecting minor changes in these outcomes is not critical for the utility of the study overall. Furthermore, it would be prohibitively expensive to conduct a study with school-level random assignment that would yield the MDES at the teacher level that are comparable to the proposed MDES for student outcomes (e.g., a MDES of 0.2 or smaller).

Steps in Sample Selection

Of the approximately 50 schools sampled in this study, half will be randomly selected to participate in the Pacific CHILD professional development program, and the remaining half will be assigned to a control group that will be excluded from the intervention for two years. The random assignment will be conducted at BPA and a BPA research staff member will notify each school of their group assignment.

An overview of school selection, recruitment preparation, and recruitment activities are listed below:

1) PREL visited the State Education Agencies to build high-level support for the study and identify potential study schools. Signed letters of interest from CNMI, American Samoa, and Hawai'i are in Appendix L.

2) PREL and BPA have started to collecting background information on the schools in each entity, including geographic location, enrollment, number of teachers, types of professional development they have engaged in, student achievement, and any other research studies the schools may be participating in. This information allows us to screen out schools that may not be suitable for the study, for example because they are too small or too remote.

3) PREL met staff from local (island-level) education agencies for preliminary information sessions. During these meetings, PREL discussed with area superintendents and school principals about the study and made presentations about the details of the study, the promise of the intervention, and the importance of the research

4) After these meetings, PREL will follow up with administrators at each level to confirm their interest and participation in the study and distributed information sheets and consent forms (see attachment I). Where necessary, PREL will make follow-up visits to meet with teachers to discuss the intervention and evaluation activities. Each principal will be asked to sign an MOU with PREL to support his or her teachers' participation in the program. A sample MOU can be found in Appendix M.

5) We make participation of any given school conditional on receiving signed letters of interest from the Commissioners of Education (CNMI and American Samoa) and signed letters from the Superintendent in Hawai'i along with signed MOUs from each participating school in Hawai'i.

2. Procedures for the Collection of Information and Estimation

Statistical Power of the Sample

<u>Considerations for a Cluster Random Assignment Design</u>

Although random assignment will take place at the school level, most outcomes will be assessed at the teacher, classroom, or student level. As a result, outcome data will be clustered within the units of randomization. This clustering reduces the study's effective statistical power because individual units of observation (teachers, classrooms, or students) are not entirely independent of one another. The interdependence of these individuals and classrooms can be assessed empirically and is expressed statistically using a parameter known as the intraclass correlation coefficient (ICC). The more interdependent individual observations are, the larger the ICC, and the smaller the effective sample size for the study.

Fortunately, the fact that observations are clustered does not in any way compromise the validity of the random assignment research design itself. As long as the units of random assignment are randomly placed in program and control groups, the resulting estimates of program effects will be valid and unbiased. The only consequence of clustering is that larger samples are needed to produce statistically significant impact estimates than would be the case if individual-level random assignment of students or teachers were a possibility.

As detailed in Schochet (2005); Bloom, Richburg-Hayes, and Rebeck Black (2005); and Bloom, Bos, and Lee (1999), strong school-level covariates are most helpful in mitigating the impact of clustering on statistical power in studies that use school-level random assignment. For example, Bloom, Bos, and Lee found that the sample sizes needed for cluster random assignment studies were generally very large when ICCs approached or exceeded 0.2. Those levels of intraclass correlation are quite common in educational settings. However, they also found that the effective ICC could be reduced dramatically by including baseline covariates in impact analyses. Even if those covariates were measured at the aggregate (school) level, including them in individual-level impact analyses reduced the effects of clustering significantly. For example, they found that adjusting for the aggregate reading test scores of a previous cohort of students reduced the effective ICC from 0.2 to 0.05 in one study of a large urban school district. They were able to reproduce this effect across several different cohorts of students and several different outcome areas, including math scores.

<u>Minimum Detectable Effect Size Calculations</u>

Our study is designed to provide reliable findings by optimizing its statistical power. Our cluster random assignment research design provides reliable and policy-relevant impact estimates for both teacher- and student-level outcomes.

Table 3 below summarizes the precision of our research design in terms of the smallest possible program effect that can be detected reliably with a certain given sample size at the 5% level of significance based on a one-tail test. The table displays the minimum detectable effect size (MDES), which is expressed as a multiple of the standard deviation of the outcomes. MDES are calculated for both student and teacher outcomes under different sample scenarios and different assumptions on the intraclass correlation level.

To calculate student-level MDES, we are using empirical data reported by Bloom et al. (2005) to determine appropriate values for the intra-class correlation (ICC) and the explanatory power of our impact regressions. The ICCs and R2 statistics reported by Bloom et al. appear to be correlated. In two school districts they found values of 0.17 and 0.77 for ICC and R^2 in one district and 0.23 and 0.91 in the other. Given these results, we believe an ICC 0.15-0.20 and an R^2 of 0.65-0.70 to be reasonable. For calculating teacher-level MDES we unfortunately have no reliable data on the explanatory power of

teacher-level covariates in impact regressions and no reliable data on teacher-level ICCs. Thus, we used an ICC of 0.10 and 0.20 to bracket a likely empirical value, and we use an R-squared of 0, which is as conservative as it can be. We also provide MDES at ICC of 0.05 assuming that regression estimation with baseline covariates will reduce the effects of clustering by lowering the intraclass correlation to 0.05. For the numbers of teachers and students per school, we assume 5 teachers and 125 students per s school based on the recent (2005-06) SEA data from the target entities. The assumption of 50 schools with 5 teachers and 125 students per school used in Exhibit 3 closely reflects the expected make-up of the sample based on the average of the schools we have recruited thus far. [1]

Exhibit 3
Minimum Detectable Effect Sizes (MDES)

MDES for student outcomes ($R^2 = 0.5$)	Num of Schools	Students per school	Total Num of Students	MDES 1 (ICC=0.05)	MDES 2 (ICC=0.1)	MDES 3 (ICC=0.2)
Full Sample	50	125	6250	0.15	0.20	0.27
Students in a subsample of 25 schools (e.g. Hawai'i schools)	25	125	3125	0.21	0.29	0.39

MDES for teacher outcomes ($R^2 = 0.0$)	Num of Schools	Teachers per school	Total Num of Teachers	MDES 1 (ICC=0.05)	MDES 2 (ICC=0.1)	MDES 3 (ICC=0.2)
Full Sample	50	5	250	0.40	0.43	0.49
Teachers in a subsample of 25 schools (e.g. Hawai'i schools)	25	5	125	0.57	0.62	0.70

Source: Calculations using Optimal Design Software (Raudenbush and Liu, 2000)
Notes: MDES 1 = MDES for ICC of 0.05, MDES 2 = MDES for ICC of 0.1. MDESs are calculated at power=0.8.

As Table 3 shows, with the target sample size of 50 schools, the design produces student-level MDES between 0.15 and 0.27 for the full sample and between 0.21 and 0.39 for students in half the schools. We believe that these MDES are sufficiently small for an ambitious and intensive intervention like Pacific CHILD. Teacher-level MDES are estimated to be larger, ranging from 0.40 to 0.49 for the full sample of teachers, using our assumptions. That is, the study will not detect a relatively small improvement in teacher outcomes, even with the full sample of teachers. However, as discussed above, we do not expect a teacher-professional development intervention like Pacific CHILD to produce impacts of 0.20 or better on students without at least a 0.40 or higher impact on teachers. Also, we believe that the student-level MDES should be the primary driver for the statistical power calculations since student-level outcomes are ultimately the primary focus of the intervention and the study.

Statistical Methodology for Stratification and Sample Selection

Nonstatistical Sampling Approach

For this school-level random assignment study, all 4th and 5th grade teachers and students of the sample schools are considered as the target populations. As described in the beginning of this document, the schools which meet certain minimum criteria are identified as the target schools for recruitment. Actual selection into the sample is a result of voluntary decisions on the part of school principals and districts, and the study team do not control their decisions. In this regards, the study sample is a convenience sample. The key criteria used to select the schools for recruitment include:

---

[1] Based on the SEA data and recruitment information from American Samoa and CNMI we have to date, we estimate that on the average, our sample will have 5.1 teacher per school and 129 students per school.

- The school cannot be in the process of restructuring as defined by NCLB standards, since these schools have very specific plans already included as part of their professional development

- The school must be geographically accessible to prevent inordinate financial and personnel resources from having to be expended to provide services and collect data.

- The school should not be implementing a professional development program or curriculum that is incompatible with Pacific CHILD.

- The school must have a minimum number of 4th and 5th grader teachers (2 per school).

- The school must have English language learner students in 4th and 5th grades.

In case of American Samoa and CNMI, the majority of schools in the entity is identified as the target schools. The total number of elementary schools is only 23 in American Samoa and 12 in CNMI. The majority of these schools are identified as the target schools. At the time of writing, 15 schools in American Samoa and 10 schools in CNMI are expected to participate in the study.

<u>Heterogeneity in the Sample</u>

Since the treatment and control schools will be randomly assigned (after blocking by entity), heterogeneity within the sample itself will not affect the validity of the study. Regardless of the extent of across-school heterogeneity, differences in average outcomes across the program and control groups will provide unbiased estimates of program impacts. However, heterogeneity within the sample of schools reduces the reliability of program impacts. This fact is reflected in our statistical power calculations discussed above, which include a significant cluster adjustment that explicitly accounts for the clustering of individual student and teacher observations within school contexts that may be very different from one another. We are confident that our planned sample size of 50 schools (and the associated sample sizes of teachers and students) is sufficiently large to produce reliable impact estimates, even if participating schools are very different from one another.

Furthermore, as discussed later, our analysis will use regression-adjusted estimates of the program impact, rather than simple differences-in-means impact estimates. The regression procedure is a key tool for minimizing the effective heterogeneity of the sample. By controlling for student-, teacher-, and school-level sample characteristics and aggregate school-level preprogram outcome variables, the procedure maximizes the share of the unexplained variation in study outcomes that is found within schools rather than between schools. This, in turn, minimizes the intraclass correlation of outcomes across schools and minimizes the effects of school-level heterogeneity on the reliability of the impact estimates. In addition, these regression adjustments minimize residual differences that arise in the observable characteristics of students and teachers in the experiment and control groups as a result of random sampling error.

Regression covariates for the impact analyses will be selected to meet the following criteria:

- They are baseline measures that pertain to the period prior to random assignment
- They have predictive power in the regression models for key outcomes
- Where possible, they include individual- and school-level lagged values of the outcome variables

On the basis of these and other criteria, the explanatory variables at all levels will be selected from the following variable categories:

Student-Level Characteristics

- Pre-random assignment test results
- Demographic characteristics, including age, sex, race/ethnicity, and primary language (may be aggregated to the level of grade or to the level of school, depending the data availability).

Teacher-Level Characteristics

- Demographic characteristics, including age, sex, race/ethnicity, and primary language
- Professional background, including years of service, education level, certification in reading instruction, and certification in ELL instruction
- Baseline information on teacher's classroom practice

School- and Jurisdictional-Level Characteristics

- Organizational characteristics, including school size, teacher-to-student ratio, average school days and instructional hours, average reading test scores, and per-student spending
- Other community characteristics, such as the neighborhood economic status (for Hawai'i schools these may include whether the school is in or outside Honolulu, and whether it is on Oahu or a neighbor island) and percentage of non-English speaking residents

As discussed in the previous section, these measures will be collected as part of the baseline data collection. Whenever possible, these measures will be collected through existing data and records. Exact explanatory measures to be included will be empirically determined based on careful examination of the quality and relevance of the data and measures, as well as the feasibility of conducting reliable analysis given the sample size.

Estimation Procedures

The description of program impacts is the central objective of the random assignment study design. A major advantage of using random assignment is that such a design does not rely on highly complex (and often controversial) data analysis techniques to produce valid impact estimates. The study design assures that post-random assignment differences between outcomes for teachers and students in the program and control groups are unbiased estimates of the program effects. However, as discussed in the section on statistical power, there are significant benefits to using baseline covariates in the impact analysis, mostly in terms of increased statistical power. Thus, we will use multiple regression models to analyze the outcome data for this study. These models will control for student, school, and teacher background characteristics, all collected prior to random assignment. To maximize the useful variance, these models will be hierarchical in nature (i.e., they explicitly acknowledge the nested nature of the data),

For covariates at the student level, we will use mostly school-level or grade-level covariates, because individual-level student background data is not always available and may lead to observations being dropped from the analysis due to missing data. For the purposes of increasing statistical power, it is most important to control for student outcomes at the aggregate (school) level. School-level covariates minimize random school-to-school variation in background characteristics between the program and control group schools. To create these aggregate covariates, we will use reading test outcomes measured prior to random assignment. At the teacher level, we will control for teacher education level, teacher experience, teacher English language ability (in non-English speaking states), as well as basic demographic characteristics such as age, gender, and ethnicity. A somewhat simplified format for a student-level impact regression model would appear as follows:

$$Y_{ijk} = \beta_0 + \beta_1 P_k + \sum \beta_z Z_i + \sum \beta_x X_j + \sum \beta_s S_k + \gamma_k + \delta_j + \varepsilon_i \quad (1)$$

In this model, Y_{ijk} represents outcome Y (a reading comprehension test score, for example), measured for student i with teacher j in school k. P_k is a program variable, which is measured at the school level and has a value of 1 for program schools and 0 for control schools. β_1 is the program effect associated with this variable. Z_i, X_j, and S_k are three vectors of control variables, for students i, teachers j, and schools k,

respectively. Each of these vectors is accompanied by a series of regression coefficients (β_z, β_x, and β_s). Separate error terms for schools, teachers, and students are represented by γ_k, δ_j, and ε_i, respectively. Although Equation 1 as written appears to represent a fixed effects regression model, we do not plan to estimate it that way, because doing so would not be appropriate given the hierarchical nature of the data. Instead, the models will be estimated as a series of three nested hierarchical models (at the school, teacher, and student levels), in which the unexplained error at one level becomes the outcome to be explained at the next level. After estimating these regression models, we will use the estimated coefficients to calculate regression-adjusted mean outcomes for program and control schools. The regression-adjusted means will be presented in tables and figures so that readers do not have to interpret regression coefficients to learn about the impacts of the program.

Information To Be Collected

Pacific Resources for Education and Learning (PREL) is the administrator of the Pacific CHILD program. PREL staff will recruit schools and teachers for the program, and provide the Pacific CHILD professional development training, support, and services. Berkeley Policy Associates (BPA) is the independent evaluator of the Pacific CHILD and is responsible for the data collection activities listed below, with the exception of the distribution of information and consent forms, which will be distributed by PREL. Of the data collection activities discussed below, we seek OMB approval for Annual Teacher Survey, Annual Impact Survey, Annual School Principal Survey, and Teacher Focus Group/Interviews.

Information Sheets and Consent Forms

PREL will distribute information sheets and consent forms to each teacher during the recruitment process.   Signed consent forms will be submitted to BPA. The information sheets, consent forms, and IRB approval letter are included in Appendix J and K.

Teachers: Each participating teacher will be provided with a consent form during recruitment. The consent form and information sheet will address all aspects of the study, including random assignment of the treatment, confidentiality, participation in the focus groups and/or interviews, surveys, and classroom observations. Teachers will be asked to sign the consent form once during the study period. BPA researchers will also provide each teacher with a separate information sheet for each data collection activity, detailing the specific procedures for that activity (see Appendix J).

Students: We will seek passive parent consent for accessing student records. The study will not directly contact students or collect any information directly from students or their parents. We will utilize the existing data maintained by the schools or districts (See the subsection titled "School or District Administrative Records" below). Passive parental consent will be sought for all students enrolled in the participating schools as 4th or 5th graders during the study period.

Pre-program Classroom Observations (Treatment and Control Groups)

BPA researchers observe regular classroom lessons of the participating teachers using the Sheltered Instruction Observation Protocol (SIOP) before the treatment starts. The purpose of the pre-program classroom observation among selected teachers is to enable researchers to determine possible variation in teaching practices across sites. Teachers will not be required to do anything different from what they would have planned on the day of the observations. The SIOP observation Protocol can be found in Appendix I.

<u>Teacher Focus Groups and/or Interviews (Treatment Group Only</u>**)**

The teachers in the Treatment Group will be asked to participate in a focus group with other participating teachers or to participate in a group or individual interview, depending on the number and availability of participants at the school. The purpose of these focus groups and interviews is to understand the teachers' experiences with Pacific CHILD and solicit teachers' feedback on the program. The focus groups and interviews will be conducted annually during the treatment years and will be guided by a discussion protocol (see Appendices E & F)

<u>Program Activity Observations (Treatment Group Only)</u>

BPA researchers will observe regular Pacific CHILD program activities, including summer training institutes, follow-up institutes, individual coaching sessions, and structured learning team meetings. The purpose of these activity observations is to document how Pacific CHILD is implemented during training and at the school.

Activity observations for training institutes will be scheduled according to the Pacific CHILD program schedule. Activity observations for at-school program activities will be scheduled annually during the treatment years. The professional development observation protocol can be found in Appendix H.

<u>Annual Teacher Survey (Treatment and Control Groups)</u>

Teachers will be asked to take a survey to assess the impact of the Pacific CHILD Program (in the case of the treatment group), and alternative professional development activities (in case the of the control group) on their instructional practices, and their outlook on teaching ELL students. This survey is expected to take place annually during the treatment years. The Teacher Survey can be found in Appendix B.

<u>Annual Impact Survey (Treatment and Control Groups)</u>

Teachers will be asked to complete a teacher knowledge survey to assess the their content knowledge and pedagogical skills. This survey is expected to take place annually during the treatment years. The Impact Survey can be found in Appendix C.

<u>Annual Program Impact Observation (Treatment and Control Groups)</u>

Teachers will be asked to allow BPA researchers to observe their regular classroom lessons once in each study year. The purpose of this program impact observation is for researchers to look for variation in teaching practices across sites after the Pacific CHILD is implemented. All teachers in the study will be observed for this outcome measure. These impact observations are expected to take place annually during the treatment years. The observation protocol can be found in Appendix I.

<u>Annual School Principal Survey (Treatment and Control Groups)</u>

School principals will be asked to take a survey about the professional development activities offered to and/or required of teachers at their schools, and about curriculum, policy, and staffing changes at their schools. If the school is in the treatment group, the survey will also ask about the implementation of Pacific CHILD. This survey is expected to take place annually during the treatment years. The Principal Survey can be found in Appendix D.

<u>School or district administrative records:</u>

School or district administrative records will be used to measure student outcomes. To minimize the burden, we will request the data transfer in two or three batches, timed according to the administration's schedule within the broader timeframe of the study. Student achievement will be measured using existing standardized test scores from (SAT 10) which are currently implemented by the study jurisdictions as part of their annual assessment. We will collect student records including standardized test results and demographic information, if available, including age, gender, race/ethnicity, ELL status, English proficiency level, special needs status, and system-specific student identification number. This information will be accessed, maintained, and analyzed according to BPA's data security protocol (see Attachment N).

Data Quality Control Procedure

In order to ensure that data collected is of high quality, we adopted several strategies. These include: choosing outcome measures and other covariates that have been shown to be reliable in previous research and/or pretests; establishing sound data collection protocols; training data collection staff adequately; closely monitoring the data collection process; and examining the data for any outliers, response errors, and abnormal patterns and making appropriate adjustments. Second, we will use multiple measures especially for outcome variables in order to check the robustness of the impact estimates to different types of instrumentation. Furthermore, these measures are used to assess student and teacher outcomes more than once during the follow-up period. Third, we will pursue statistical means, such as structural equation modeling, to examine and address the reliability of key outcome measures if we suspect serious measurement errors in our variables.

3.   Methods for Maximizing Response Rates

PREL and BPA will implement a number of different strategies to reduce participant burden and to increase participant buy-in:

- We will make comparable professional development services available to control schools and teachers after a 2-year embargo period. This makes participation in the study less of an obvious burden for control schools, which otherwise might feel that they do not benefit from the study.

- We will use standardized mandated achievement tests as student outcome data, so that there will no additional test-administration burden on teachers and school administrators.

In addition to securing buy-in to study by target schools through careful recruitment, it is important to make participation in the study attractive for schools and teachers, and to sustain participation of both treatment and control groups. It is particularly important in a school-based random assignment study to avoid attrition and crossover problems at the school level, since the loss of even a few schools can greatly reduce statistical power. We will conduct high-level, in-person recruitment visits to each jurisdiction to obtain buy-in to the study and to select schools that have stable conditions. As noted, control schools will be offered delayed access to the intervention. Furthermore, we propose that:

- PREL and BPA staff will discuss the importance of the study for regional/national education policy and the benefits of the study for the jurisdiction and its students. We will invite local school principals and other immediate stakeholders to initial meetings to ensure that buy-in is not limited to district-level management. Follow-up meetings will be held at schools selected for the study. These meetings will include principals, relevant teachers, and parents. These meetings will emphasize the ultimate benefits of the study for both treatment and control schools, teachers, and

students.

- We will produce easy-to-read and well-designed information materials about the study, its goals and regional importance. This information will target parents, teachers, and students, providing them with a clear rationale for the research project and seeking to enlist them as advocates for the study.

- Through the implementation monitors, we will conduct careful periodic assessments of study implementation in all the participating districts. These assessments should be followed by constructive discussions with school and district staff to reinforce their continued commitment to the project and help them troubleshoot any concerns.

- We will use Memoranda of Understanding (MOUs) (see Appendix M) and other contractual arrangements to create a strong framework of agreements and mutual obligations that make it difficult for any participating school or district to drop out of the study.

At the individual participant level, we have designed the study to achieve a response rate of at least 80 percent, based on our past experience in collecting data in school settings as well as in various experimental and quasi-experimental studies that required multi-year follow-up. In our experience, it is most important to carefully train data collection staff, to closely monitor the progress of data collection, and to make quick and decisive adjustments to protocols when response rates are at risk of falling below key targets. Data collection staff will be carefully trained to: 1) maintain up-to-date contact information for all study participants, including multiple forms of contact; 2) contact and collect data from students and teachers who are absent from scheduled data collection sessions; 3) track responses in an integrated project data system; 4) contact PREL/BPA staff immediately regarding lower than expected response rates, crossovers, or other difficulties. Implementation monitors will also assist in identification and reporting of crossovers. Senior research and project staff will meet monthly to discuss any problems and revise protocols or intervene with sites as needed.

At this time, we are not planning any specific statistical adjustment to address bias due to potential sample attrition. Our primary plan is to keep the sample attrition to a minimum, and we will attempt to follow up with teachers and students if it is feasible. Also, any effect of attrition- due to normally expected transfers of teachers and students- on the impact estimates could indeed be considered as part of the policy-relevant impact estimation, rather than a bias to be corrected. In any case, we will monitor the sample and document the circumstances of attrition and any likely implications for program implementation and outcomes. The detailed documentation will allow us to assess whether and what type of statistical adjustments may be necessary and appropriate when it comes to the time of estimation.

4. Test Procedures

All new data collection instruments were tested with 9 or fewer teachers in three jurisdictions in the Pacific. Because of the small sample size, item level statistical analysis was not conducted. However, changes were made to improve the clarity of survey/test items and increase the relevance of observation protocol items. Testing did not lead to significant changes in the length or content of instruments. The time burden estimate associated with each instrument was obtained through this piloting.

The SIOP© observation protocol (Pearson Education, Inc., 2004) will be used for classroom observations. This instrument has been well validated by the developers in a variety of settings with ELLs (Echevarria, Vogt and Short, 2004).

5.   Names of Statistical Consultants

Technical Working Group Members Who are Consulted on Statistical Aspects of the Design

Name	Affiliation	Telephone	Email
Dr. Geoffrey Borman	Interdisciplinary Training Program in the Education Science, Univ. of Wisconsin-Madison	(608) 263-3688	gborman@education.wisc.edu
Dr. Robert Boruch	Graduate School of Education, University of Pennsylvania	(215) 898-0409	robertb@gse.upenn.edu
Dr. Thomas Cook	Dept. of Sociology Northwestern University	(847) 491-3776	t-cook@northwestern.edu
Dr. Hiro Yoshikawa	Graduate School of Education Harvard University	(617) 384-8250	hiro_yoshikawa@havard.edu
Dr. Shuquiang Zhang	College of Education University of Hawai'i	(808) 956-4295	szhang@hawaii.edu

Study Team (Responsible for data collection and analysis)

Name	Affiliation	Telephone	Email
Dr. Roger Chesswas	Pacific Resources for Education and Learning (PREL) 900 Fort Street Mall, Suite 1300 Honolulu, Hawai`i 96813	(808) 441-1300	chesswasr@prel.org
Dr. Johannes Bos	Berkeley Policy Associates (BPA) 440 Grand Ave, Suite 500 Oakland, CA 94610	(510) 465-7884	hans@bpacal.com
Dr. Yasuyo Abe	Berkeley Policy Associates (BPA) 440 Grand Ave, Suite 500 Oakland, CA 94610	(510) 465-7884	yasuyo@bpacal.com

References:

Schochet, P. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research.

Bloom, H. S., Richburg-Hayes, L., and Rebeck Black, A. (2005). Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions. New York: MDRC.

Bloom, H., Bos, J., & Lee, S. (1999). Using cluster random assignment to measure program impacts. *Evaluation Review*, 23(4), 445–469.

Bloom, H. S., Bos, J. M., Gennetian, L., Hill, C., Michalopoulos, C., Morris, P. et al. (2005). *Learning more from social experiments: Evolving analytical approaches*. New York: Russel Sage Foundation Press.

Echevarría J., Vogt, M. E., & Short D. J. (2004) *Making content comprehensible for English language learners: The SIOP model* (2nd ed.). Boston: Allyn & Bacon.

Raudenbush, S.W., J. Spybrook, X.,Liu, and R. Congdon. (2004). *Optimal design for longitudinal and multivariate research: Documentation for the "Optimal Design" software*. Working paper.