Impact Study: *Understanding Science* Professional Development and the Science Achievement of English Learners

CONTRACT NO. ED-06-CO-0014

Request for OMB Approval

Supporting Statement Part B

Submitted by:

REL WEST

&

Heller Research Associates

November 15, 2007

Table of Contents

List of Tables

Section B: Collection of Information

B1. Respondent Universe and Sampling Procedures

The unit of assignment in this study is teachers who will be randomized to treatment and control condition within schools. Based on our power analysis, we are planning to recruit about 120 teachers and 4,800 students from at least 6 research sites. The respondent universe for this study is 8th grade middle school teachers in California public schools (specifically urban schools with high proportions of English learners) as well as the 8th grade students within these schools. However, this sample is not a probability sample, and is therefore not intended to represent this population. Sampling procedures begin with a multi-layered recruitment process that determines which districts the respondents will be drawn from.

Recruitment and Assignment to Condition

Research sites will be identified and recruited by the WestEd *Understanding Science* project team, which is responsible for delivering the staff development course. The *Understanding Science* project team has existing networks of sites and science teachers, which include middle-school teachers, as a result of other science projects and development work currently funded by National Science Foundation (NSF). At least 6 regional sites will be identified, from each of which teachers will be recruited through local site coordinators. Details of the recruitment and sampling procedures are as follows.

Securing formal agreements with districts. The recruitment process begins by contacting county and district science coordinators who are likely to consider conducting a large-scale study on science professional development in their district. Interested science coordinators arrange for the recruitment team to meet with science staff development staff (e.g., teacher leaders and staff developers). Discussions focus on identifying individual districts or consortia of districts that (a) have an existing infrastructure for providing a program of science staff development; (b) include a core of staff developers who have the science background, inclination, and time to facilitate the *Understanding Science* courses; (c) are willing to participate in a study focusing on science and that involves random assignment of teachers; (d) have a proven ability to recruit science teacher participants in sufficient numbers for the study; and (e) are able to bring teachers together in a central location for an ongoing series of staff development sessions. If these exploratory conversations generate mutual interest and good fit, the recruitment team contacts the superintendent, in coordination with the district science coordinator.

Once oral confirmation of study participation is received, a memorandum of agreement (MOA) will be sent to each site outlining what support sites will receive for participating in the study, the roles and responsibilities of both research staff and site staff, and estimates of the time required to collect data (see a sample of MOA in Appendix I).

Recruiting site coordinators. Once districts have formally committed to participate in the project, site coordinators are chosen with input from district leadership and project staff. Coordinators then attend a two-day meeting in Oakland, CA to learn more about the study, experience components of the staff development course, and gain familiarity with their roles and responsibilities (e.g., help recruit teachers and facilitators, ensure the experiment runs as planned, and troubleshoot problems as needed).

Recruiting facilitators. Upon returning to their districts, site coordinators identify and solicit the participation of staff development leaders who have at least two years experience leading staff development in middle school science. Members of the recruitment team choose facilitators from individuals who show interests in participating in the study, with input from district leadership. Facilitators will be trained to lead the professional development course in pairs (i.e., two facilitators per site).

Recruiting teachers. Because teachers are required to teach force and motion in 8^{th} grade by the California state standards and district mandates, teachers at this grade will be recruited.

To qualify for participation, teachers must (a) currently teach 8^{th} grade physical science; (b) have taught for at least one year prior to participating in the study; (c) agree to be randomly assigned to either the treatment group, to attend an *Understanding Force and Motion* course in August 2008, or the delayed-staff-development control group, to attend the course in August 2009; (d) teach a classroom force and motion unit in fall 2008, planning to complete it by December 31, 2008; (e) provide teacher, student, and classroom data for the course evaluation. Previous recruiting experience has indicated that on average, approximately 5% of teachers solicited actually fit the criteria and profile that both qualify for participation and have the time, schedule, and flexibility to permit their involvement in a randomized study.

Coordinators at each site will recruit at least 20 teachers, half of whom will later be randomly assigned to the treatment group and half to the control. Recruitment brochures with information such as that in Appendix A, PowerPoint presentations, and a web site will be used to streamline the recruitment process and ensure that teachers receive accurate and consistent information about the study, professional development, and roles and responsibilities of participants in the study. As a condition for participation, teachers must agree to complete the force and motion unit at least one week prior to the administration of state standardized tests in order to provide enough time to administer the posttests before statewide testing begins.

Teachers in both groups may take part in any other science professional development during the school year. Teachers who are randomly assigned to the control group will be informed that they will be guaranteed placement in the course during summer 2009. The teachers in the control group will be free to take part in any other science professional development that arises during the school year but will receive no additional treatment as part of this study. Assessment measures will be administered to control groups in each district at the same time that the treatment group measures occur.

Selecting classes of students. Teachers in both conditions will be asked to collect data from two of their eighth-grade physical science classes that meet the following criteria: (a) class is one of the teacher's first two classes in the morning; and (b) at least one class is classified as mainstream students, not exclusively accelerated or special needs. In this way the teachers' choices will be partially constrained, so as to prevent teachers from selecting only their strongest class sections.

B2. Statistical Methods for Sample Selection and Degree of Accuracy Needed

Stratification and Sample Selection

Teachers will be randomly assigned within schools. Thus, schools will serve as strata for the random assignment. See section B1 for the discussion of sample selection.

Degree of Accuracy Needed

In order to determine the appropriate sample sizes required for the experimental study, we calculated minimum detectible effect sizes (MDES) based on the unit of randomization, the sources of clustering, the availability of baseline explanatory variables, and other design characteristics using the procedures described by Donner and Klar (2000), Murray (1998), Raudenbush (1997), and Schochet (2005). MDES estimates represent the smallest true program impacts in standard deviation units that can be detected with high probability (Bloom, 1995). As defined in our design work, the MDES of a particular study is the smallest effect size that has at least an 80% probability of being found statistically significant with 95% confidence. For a design to be sufficiently powerful, this MDES must be small enough so that a likely program impact that is large enough to be policy- relevant does not go undetected.

As discussed above, 120 teachers will be randomly assigned to two conditions. We assume that each teacher will cover two sections with approximately 25 students per class. We conservatively assume a student non-response rate of about 20% for power estimation purposes, leaving 20 students per class and 40 students per teacher at the end of the spring semester for analysis. For the purposes of the power analyses, we conservatively assume intraclass correlations (ICCs) of 0.20 for the student academic outcomes. Although Schochet's (2005) recent work suggested ICCs closer to 0.15, recent work by Bloom, Richburg-Hayes, & Black (2006) and Hedges & Hedberg (2006) suggest values closer to 0.20 for middle school students. To incorporate uncertainty about the level of ICCs – we present MDES estimates for a range of ICCs in Table B1. We present MDES estimates for two sets of assumptions with regard to the explanatory power of covariates – assuming (a) between- and within-teacher R^2 values of .50 based on Schochet's (2005) work, and (b) a between-teacher R^2 value of 0.75 and within teacher R^2 value of 0.56 based on Bloom et al.'s (2006) work. However, because Bloom et al.'s analyses are applicable to designs in which schools rather than teachers are the unit of assignment, and may not be generalizable, we think that assumption (a) is more appropriate. We conservatively assume that covariates will explain 20 percent of the variance in teacher outcomes based on work by others focusing on teacher outcomes (e.g., Hill and Ball, 2004; Schweingruber & Nease, 2000).

Table B1. Minimum Detectable Effect Sizes (MDES) for Different Combinations of Teachers, ICCs, and Within- and Between-Class Explained Variance

				R^2_I = .50 & R^2_C = .50 Intraclass Correlation			R^2_I = .56 & R^2_C = .75 Intraclass Correlation		
Schools	Teachers	MDES	Students	r=.10	r=.15	r=.20	r=.10	r=.15	r=.20
30	60	.66	2400	.18	.22	.24	.14	.16	.18
40	80	.57	3200	.16	.19	.21	.12	.14	.15
50	100	.51	4000	.14	.17	.19	.11	.12	.14
60	120	.46	4800	.13	.15	.17	.10	.11	.13
70	140	.43	5600	.12	.14	.16	.09	.10	.12
80	160	.40	6400	.11	.13	.15	.08	.10	.11

Notes: R^2_I and R^2_C refer to the proportion of within teacher/classroom and between teacher/classroom variance explained, respectively. We assume that each teacher covers two class sections with approximately 20 students with valid outcome data per section.

Shaded cells correspond to expected MDES.

Calculations are based on the following assumptions: 1) equal numbers of teachers assigned to experimental and control conditions, 2) statistical power levels of .80, 3) Type I error rates of .05 (two-sided), 4) a fixed-effects statistical model, and 5) covariates used in the analysis explain 20% of the variance for teacher outcomes.

With 60 teachers per condition and a minimum of 40 (25*2*.80) students per teacher, we estimate the MDES to be 0.17 for student academic outcomes. With as few as 7 students per teacher, the MDES only rises to 0.20 – suggesting that adequate power is available for conducting analyses of student subgroups. However, statistical power for estimated differences in impacts between student subgroups (research question #2, as indicated in A1) is not high. With 60 teachers per condition, 40 students per teacher, and an ICC of 0.20, the estimated minimum detectable "difference in the difference" interaction effect size is 0.24 standard deviations, which is a substantial difference relative to a main effect of 0.17 standard deviations.

The MDES for teacher outcomes is 0.46. Teacher level MDESs of such magnitude are acceptable because impacts at the more proximal teacher level will tend to produce smaller subsequent impacts at the more distal student level.

B3. Methods to Maximize Response Rates and Deal with Nonresponse

A high level of sample attrition is unacceptable for the integrity of the experimental design. Sample attrition relates to our ability to collect outcome data on all teachers who were randomly assigned at the start of the study. Serious violations in this regard will likely cause significant biases in the estimated program effects. For this reason, it is critical that any teachers who agree to participate in this study remain involved in the research efforts until all data collection is completed, even if they were unable to fully participate in the intended professional development treatment. This is a key focus of our upfront recruitment efforts.

Multiple methods will be employed to maximize response rates (at least 80%), including strong mutually beneficial relationships, good survey design, complete collection of contact information, and persistent follow-up. Survey data will be processed immediately, and non-respondents will be scheduled for follow-up administration[1]. In terms of response rates, keeping the burden to a minimum is, of course, the first device for ensuring adequate response rates, but we recognize that that alone is often insufficient.

To establish good relationships with the project staff, communication with teachers will be both friendly and efficient. After applying and being accepted for participation, teachers will receive welcome email messages that give them essential information about the course and data collection (see Welcome Letter for Teachers in Appendix J). They will be provided with access to a web site that provides all necessary information about their responsibilities, kinds of data to submit and dates of each data collection activity. This web site will also provide teachers with individualized information as to which data have been received from them and what is outstanding.

Although this study includes a plan to monitor and ensure implementation fidelity, it is possible that some participants assigned to the treatment group will not participate in all intervention activities. Nonparticipation by significant numbers of those targeted to receive the intervention would likely dilute potential program impacts. Extensive efforts will be made to collect data from such non-participants, and levels of participation in the intervention will be monitored through surveys and records. So as not to bias impact estimates, all such participants will be kept in the impact analysis in their original, assigned groups to avoid sample selection bias. That is, an intention-to-treat analysis (ITT) will be performed. ITT refers to the fact that random assignment only establishes an "intention to treat," but does not actually guarantee that those assigned to the program experience it.

B4. Test of Procedures and Methods to be Undertaken

We conducted a pilot of the proposed study with a group of nine middle-school teachers in the San Francisco Bay Area during the 2006-07 school year. The teachers took the *Understanding Force and Motion* course in June 2006, and completed teacher surveys, science tests, and interviews before and after the course, as well as after they taught the force and motion unit in their classrooms. Student tests of force and motion were administered by each teacher within two weeks before and after the classroom unit.

[1] Three attempts via either email or phone call on a bi-weekly basis.

For the most part, the questions on surveys used in the pilot were those that we have used in several previous studies, so they are known to be reliable and comprehensible to respondents. All force and motion tests had been used in other studies and have been shown to possess good psychometric properties. New survey questions specific to this course were piloted to make sure that the content of the questions is appropriate and the wording clear. In addition, the teacher interview focused on science knowledge for teaching was piloted pre and post-professional development because it contained newly written scenarios and samples of study work that were drafted for this study. Based on the teachers' responses, as well as interviews with the teachers about the interview questions themselves, the Teacher Science-for-Teaching Interview went through several iterations until it reached the version included in Appendix D7.

B5. Statistical Consultants

Name	Title	Responsibilities	Phone Number
Tom Hanson	Senior Research Associate, WestEd	Oversee data analysis	(562) 799-5170
Sophia Rabe-Hesketh	Professor of Educational Statistics, UC Berkeley	Consult on HLM data analysis	(510) 642-5287

Thomas L. Hanson, PhD, is a Senior Research Associate in the Health and Human Development Program at WestEd and Co-Director of research of WestEd's Regional Educational Laboratory (West) (REL West). He directs the *Lessons in Character Outcome Evaluation* (REL West/Ed-IES) and the *Tribes Outcome Evaluation* (NIJ). Hanson also serves as lead methodologist for the *Algebraic Interventions for Measured Achievement* (ED/IES), an experimental trial testing the efficacy of an intervention curriculum targeting specific algebraic learning trouble spots; *Math Pathways and Pitfalls Lessons for K-7 Students* (ED/IES), a cluster-randomized trial investigating the efficacy of the Math Pathways and Pitfalls instructional materials on 4^{th}-6^{th} grade students' mathematics achievement and mathematical language development; and the *Integrating Literacy and Science Instruction in High School Biology Project* (NSF) and *Efficacy of Reading Apprenticeship Professional Development for High School History and Science Teaching and Learning* (ED-IES) studies, which are cluster-randomized trials that examine the effectiveness of teacher training in the integration of reading instruction and subject area content on student achievement in science, history, and reading.

Sophia Rabe-Hesketh, PhD, Professor of Educational Statistics at UC Berkeley, is a statistician conducting methodological research in multilevel and latent variable modeling. She has developed a modeling framework called GLLAMM (Generalized Linear Latent and Mixed Modeling) and written a publicly available software package called gllamm (http://www.gllamm.org/) to estimate these models. Her recent books include *Generalized Latent Variable Modeling* and *A Handbook of Statistical Analyses using Stata* (3rd edition), both published by Chapman & Hall/CRC. Recent and forthcoming papers include "Generalized Multilevel Structural Equation Modeling," in *Psychometrika* (with A. Skrondal et al., 2004); "Maximum Likelihood Estimation of

Limited and Discrete Dependent Variable Models with Nested Random Effects," in *Journal of Econometrics* (with A. Skrondal et al., 2004); "Multilevel Logistic Regression for Polytomous Data and Rankings" in *Psychometrika* (with A. Skrondal, 2003); and "Parameterization of Multivariate Random Effects Models for Categorical Data" in *Biometrics* (with Skrondal, 2001). She is holding research courses at several international conferences including the 2004 Joint Statistical Meetings in Toronto. She is also involved in collaborative projects with researchers in education and psychiatry as reflected by approximately forty articles in nonstatistical journals.