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 A

Submitted by:

REL WEST

&

Heller Research Associates

March 18, 2008

Table of Contents

List of Tables

List of Figures

Section A: Justification

A1. Circumstances That Make Data Collection Necessary

This information collection is being conducted as one of the Task 2 Studies (Rigorous Applied Research and Development) of the 2005-2010 Regional Education Laboratories Program. The current authorization for the Regional Educational Laboratories program is under the Education Sciences Reform Act of 2002, Part D, Section 174, (20 U.S.C. 9564), administered by the Institute of Education Sciences' National Center for Education Evaluation and Regional Assistance. WestEd and its partner, Heller Research Associates (HRA), are collecting these data to evaluate the Understanding Science Professional Development Program for the Institute of Education Sciences (IES) of the US Department of Education.

Importance of the study

The No Child Left Behind Act of 2001 calls for all teachers to be highly qualified by the end of the 2005–2006 school year, "demonstrat[ing] competence in all the academic subjects in which the teacher teaches" (NCLB, 2002, Sect. 901(23)). This requirement is supported by evidence that a critical factor determining student achievement is teacher quality (Collias, Pajak, & Rigden, 2000; Ferguson & Ladd, 1996; Rivers & Sanders, 2002; Sanders & Rivers, 1996). In particular, grade 6-8 teacher effects are a dominant factor affecting student academic gain (National Research Council, 2000; Wright, Horn, & Sanders, 1997) and science teachers' minimal background is especially profound (Broughman & Rollefson, 2000; Fulp, 2002; National Center for Educational Statistics, 2002). As an example, two-thirds of grade 6-8 teachers received their undergraduate degrees in areas other than science or science education (Fulp, 2002). In addition, "out-of-field" teaching is widespread and a large number of middle-school teachers "move up" from elementary classrooms, placing significant numbers of individuals in middle school classrooms without training in the discipline they are being asked to teach. Inadequacies in teachers' science knowledge and pedagogical content knowledge are particularly clear in the specific impacts on their students (Brewer & Goldhaber, 2000; Monk, 1994; Monk & King, 1994; National Research Council, 2000; Rowan, Chiang, & Miller, 1997). It is therefore not surprising that students' academic performance in science has suffered, especially among underserved groups (NAEP 2006). Taken together, this research suggests improving teacher content knowledge as an important leverage point for addressing poor student achievement in science.

Unfortunately, despite the dedication of considerable resources, the great majority of teacher education and staff development programs do not result in significant gains in teachers' science content knowledge, nor do they bring about meaningful changes in teachers' instruction or assist them with the specific needs of teaching English learners (Elmore, 1996; Gandara, Maxwell-Jolly, & Driscoll, 2005; Weiss et al., 1999, 2001). The demand for academic language competencies increases as students move through grades K-8, with the science curricula becoming more rigorous and the language more complex. Students need help acquiring and extending complex comprehension processes that underlie skilled reading in the subject areas (Greenleaf, Schoenbach, Cziko, & Mueller, 2001). Currently, the science achievement of students who are learning English lags well behind native English speakers in the U.S. (Torres & Zeidler, 2002). If staff development is to address current inadequacies in middle-school science teaching, courses must address a range of teacher needs, including stronger knowledge of the content, more sophisticated pedagogical content knowledge, and specific support for working with students around issues of literacy and language in science (Fulp, 2002).

The goal of the proposed study is to investigate how to prepare middle-school teachers to improve all students' physical science content knowledge, including that of low-performing students and English learners (ELs). The *Understanding Science* courses developed by WestEd have shown promise for increasing teacher content knowledge and student achievement in national field tests. The logic model motivating this approach (see Figure 1) describes the cascade of influences connecting teachers' experiences in *Understanding Science* staff development to student outcomes. The theory of action underlying the approach posits that when the staff development is situated in an environment of collaborative inquiry—one that is rich in talk about scientific meanings in conjunction with a focus on student thinking and critical analysis of practice—this leads to increases in teachers' science content knowledge and pedagogical content knowledge, along with the acquisition of targeted strategies for eliciting and revealing student ideas, and greater skill in developing students' science language abilities. These outcomes for teachers in turn are theorized to result in changes in classroom practices, such as increased accuracy of science representations and explanations, a focus on helping students develop conceptual understanding and assessing student ideas during instruction, and a greater opportunity for students to develop the language of science. These classroom changes are hypothesized ultimately to produce improvements in student achievement, along with increased development of all students' academic language in science, and reduced achievement gaps for low-performing students and ELs.

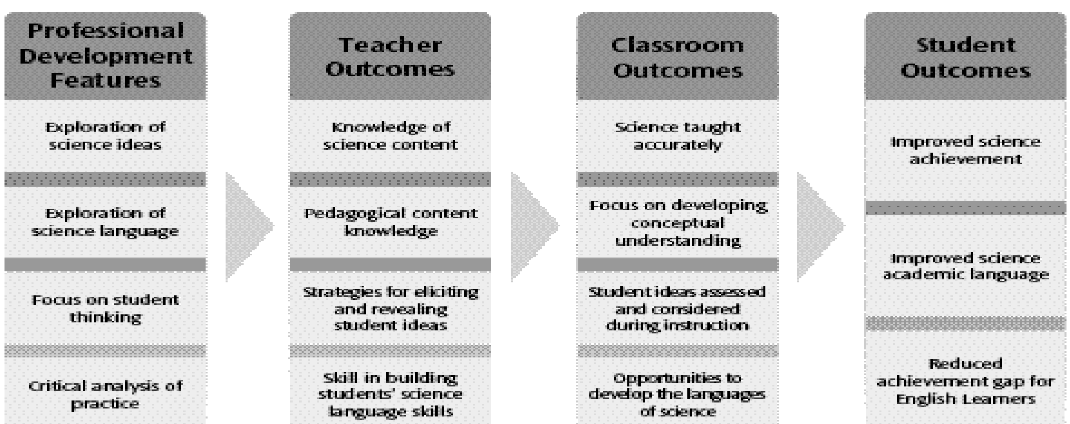

Figure 1. Understanding Science logic model showing links between intended outcomes and features of the staff development.

Research Questions

The premise underlying this work is that, to improve students' science achievement, teachers must themselves have accurate and sound science content knowledge, as well as strategies for teaching the particular science content to increasingly diverse student populations. This requires teacher education and staff development that contextualize strategies for teaching in subject-specific knowledge about: (a) the science content, here with respect to force and motion; (b) student thinking about and difficulties learning that content; and (c) instructional strategies for supporting science learning and academic language development of all students, including ELs.

The following research questions will be addressed:

1. What is the impact of *Understanding Science* staff development courses on teachers' science content knowledge?
2. How effective are *Understanding Science* staff development courses for improving students' science achievement?
    a. What is the impact of teachers' participation in the courses on their students' content knowledge in science?
    b. How effective is the program for reducing the science achievement gap between low-performing and higher-achieving students and between ELs and English-proficient speakers?
3. In the classrooms of teachers participating in the course, what kinds of changes, if any, occur in the (a) accuracy and soundness of science content communicated to students, (b) opportunities provided for student discussions about science, and (c) opportunities for students' academic language development.

Utility of Information Collected

The staff development course being evaluated in the study is designed to help teachers learn science content and pedagogical content knowledge, which in turn is meant to improve their teaching. The materials have the potential to significantly improve the methods for teaching science and pedagogical content knowledge to novice and experienced teachers.

The ultimate long-term contribution of this proposed work is greater potential for nation-wide gains in middle-school students' science achievement, resulting from widely available, low-cost staff development courses that enhance teachers' science content knowledge and improve their teaching practices.

More specifically, middle school science teachers who participate in the proposed evaluation study will receive 24 hours of staff development facilitated by district staff developers. Based on previous evaluation results, this staff development experience is in turn expected to help teachers:

- Learn major concepts of grade 6-8 physical science
- Examine how children make sense of those concepts
- Analyze and improve their science teaching

The staff development supports teachers' *classroom practices* in science so that:

- Science is taught accurately
- There is a focus on developing conceptual understanding
- There is a focus on developing academic language
- Students have increased opportunities to talk to learn in science

Again, based on previous evaluation results, anticipated benefits to *students* include:

- Improved science achievement
- Improved academic language
- Reduced achievement gaps for ELs

While these benefits will not necessarily accrue to every teacher and student in the study, the probability that these participants will benefit is high, and the risks to them of participating in learning experiences are negligible. The districts that participate in this research will serve a primarily urban student population and are home to many traditionally underserved groups. ELs comprise 25% of California students. *Understanding Science* courses were designed with features that are particularly well adapted for low-performing students and ELs. In previous studies on *Understanding Science* courses, ELs of treatment teachers made statistically significant pre-post gains in science, and science achievement of English learner students of teachers in the treatment group was significantly higher than that of English learner students in the comparison group.

A2. Purpose and Uses of the Data

In order to answer the above research questions, this study will use a cluster-randomized experimental design to test the effectiveness of WestEd's *Understanding Science* model of staff development (as shown in Figure 1), an approach that incorporates science content, analysis of student work and thinking, and critical analysis of issues related to teaching that content to students. The staff development course sessions focus on science concepts both in the context of structured investigations and in narrative cases of teaching practice drawn from actual classroom episodes involving those concepts.

This model will be evaluated by comparing it with a control condition that provides no additional science staff development beyond that already received in each school. The experiment will evaluate the value added for grade 8 teachers in California who take an *Understanding Science* course in addition to whatever science staff development they ordinarily receive. The ultimate outcome of interest is the impact of the staff development on students' science achievement, so several measures of student science content knowledge and academic language will be measured and analyzed**.** To provide a basis for explaining the results, impacts will also be studied on teachers' science content knowledge, and a descriptive study will examine selected aspects of classroom science instructional practices. Detailed research design and data collection procedure and timeline are presented below.

Research Design Overview and Timeline

Table A1 below provide an overview of the study design and measurement schedule. The study will run in California from spring 2008 through spring 2009. Outcomes will be measured for teachers, students, and classrooms through data collected during both the 2007/08 and 2008/09 school years. Course facilitators will be trained in summer 2008. Subsequently teachers in the treatment group will take an *Understanding Force and Motion* course in August 2008. Teacher pre- and post-course outcome measures will be administered in the spring before and the spring after the staff development courses are run. The spring 2009 teacher data will be collected after they have taught the force and motion unit in winter 2008-09. Students will take a force and motion pre-test and post-test within two weeks before and after a classroom instructional unit on force and motion that can occur in either fall or spring during the 2008-09 school year. These students will have taken California standardized science tests in spring 2008 and will do so again in 2009, and their scores from both years will be obtained. Classroom observation and audio data on instructional discourse, along with teacher interviews about the observed lessons, will be collected in a sample of 24 participating teachers' classrooms in spring 2008, during the force and motion unit taught prior to the intervention, and again during the 2008-09 school year.

Table A1. Overview of *Understanding Force and Motion* Experimental Design

		2007-08	**2008-09**		
	Randomization	Spring	Summer	Fall	Spring
Teachers					
Treatment	R	O	**PD**		O
Control	R	O			O
Students					
Treatment	No	O^A		O^B **FM*** O^B	O^A
Control	No	O^A		O^B **FM*** O^B	O^A

O = Observations or measurement points
PD = *Understanding Science* PD
FM = Student unit on force and motion
O^A Student standardized tests in science
O^B Project-administered student science assessments

*This can occur in spring 2009 if necessary.

Assignment to Intervention Condition

Each district will have a coordinator who will assist project staff in recruiting course facilitators and teacher participants, and in running the study at that location. All 8th-grade physical science teachers in each school (typically 2-3) will be invited to participate in the study through distribution of information about the study in flyers and on a web site (see sample teacher recruitment materials in Appendix A). Participating teachers will be randomly assigned within schools to either the treatment or control group and will receive stipends and continuing education credits for their involvement. Students will not be randomly assigned to teachers, but rather will be in their normally assigned classes.

At each of six research sites, groups of approximately 10 teachers will participate in staff development courses run simultaneously during August 2008. In addition, there will be 10 teachers in the control group at each site for a total sample of approximately 120 middle-school teachers. Outcomes will also be measured for students in two sections of physical science per teacher, for a total of approximately 4,800 students (assuming 20 students with both pre-instruction and post-instruction assessments, in each class). We anticipate that approximately one-third of the students will be ELs.

Crossovers

One of the serious challenges of this design in which the teachers are the unit of assignment within schools is that the close proximity of treatment and control teachers increases the possibility of control group contamination. This is particularly true at the middle-school level, at which teachers typically work in subject-area and grade-level teams that make detailed group decisions about curricula and instruction. For example, there is the potential for control teachers to learn about the content and approaches of the *Understanding Science* course, and even to look at the binder of materials from the course. Or the treatment teachers, who typically get quite excited about *Understanding Science* courses, could spontaneously share their newfound content knowledge or pedagogical strategies with their colleagues when they plan the force-and-motion unit.

Various steps will be taken to prevent and correct breaches in the planned random assignment. For example, rather than assuming that contamination does not occur, our plan is to respectfully inform the teachers to enlist their intentional cooperation. The aim is to elicit the teachers' collaboration in maintaining the integrity of the random assignment through building an understanding of, and commitment to, the research. In addition, agreements regarding the integrity of the design will be signed by all participants (see sample Teacher Agreement to Protect Study in Appendix B).

Contamination effects from treatment crossover will be monitored through periodic site check-in phone calls/emails/visits and the inclusion of questions on surveys asking about exposure to specific *Understanding Science* materials, activities, and principles, and about changes unrelated to the interventions. Such changes include exceptional events, changes in policies, changes in leadership, and changes in local site environments.

It should be noted that if there were contamination, this would lead to a decrease in the treatment effect size because it would reduce the distinctions between the groups. That is, the danger is that real differences will not be detected. Therefore, if significant differences were observed between the groups, this would be despite contamination, not because of it. If anything, the results would be underestimating the true effect size because of the contamination, making our conclusions stronger as to course effects.

Intervention Strategy

The staff development intervention consists of a 24-hour force and motion course for teachers, composed of eight three-hour sessions held in August 2008 over a period of five days. The eight sessions are sequenced such that the science topics (e.g., speed, velocity, acceleration, balanced and unbalanced forces) build on each other, and the corresponding science language issues and strategies for supporting student learning and language development are unveiled incrementally throughout the sessions.

There are two main components of each three-hour session—science investigations and case discussions. Science investigations engage teachers in the teaching case's core dilemma. These hands-on experiences parallel those of students in the cases, in the context of commonly-used, standards-based curricula. Following the science investigation, teachers examine student thinking and critically analyze instruction presented in the case. Case discussions lead teachers to explore alternative perspectives and solutions, which in turn causes them to rethink their instruction. Language and literacy course components teach teachers how to more effectively support students' science talk, both to make sense of the science and to help students, particularly English learners, develop their academic language proficiency (see *Understanding Science* Literacy Framework in Appendix C).

Facilitator Training. In July 2008, six pairs of facilitators from each site will be trained to lead the course in one five-day event held at WestEd in Oakland, CA. Facilitators will experience the staff development treatment themselves and be introduced to the purpose and design of the research. The majority of the training time will be spent deepening facilitators' understanding of force and motion, grounding them in the common yet incorrect ideas students (and adults) have about the science, and helping participants develop the necessary facilitation skills. Project staff will model facilitation, engage the group in analyzing video clips of exemplary facilitation, and provide practice sessions in which the trainees facilitate course sessions.

Course Materials. To ensure consistency of the treatment across sites, materials for the course will be made available as a set of two books—a Participant Book for teachers and a Facilitator Guide for staff developers.

Participant Books contain eight chapters (one per session) and present the materials teachers need to participate in all eight sessions of a staff development course. Each chapter contains a *teaching case* of actual classroom practice that illustrates students' science thinking and highlights an important teaching dilemma that any teacher might face, along with a companion *content guide* that explains and illustrates core science concepts, plus *science investigation handouts* and *case discussion handouts* that guide teachers small group working time and structure their conversations around science, student thinking and instruction.

Facilitator Guides also contain eight core chapters (one per session) and provide extensive support materials and detailed procedures needed to successfully lead a course. Each chapter describes the underlying science (including common yet incorrect ideas of children and adults) and provides scripted yet flexible procedures, such as instructions to guide the hands-on and sense making work in each science investigation, guiding questions for each case discussion, and instructions for completing classroom connection assignments teachers do between sessions.

Treatment Group. In August 2008, teachers who are randomly assigned to the treatment group will take the course *Understanding Force and Motion*, led by pairs of trained facilitators at each district, one of whom will be selected as the lead facilitator and the other who will co-facilitate and serve as the backup in case the lead facilitator leaves the study. Districts will help identify and solicit the participation of staff development leaders who have at least two years experience leading teacher staff development courses in middle school science.

During the 2008-2009 academic year, participating teachers will teach their force and motion unit when they ordinarily would (fall or spring), using their district curriculum or other materials as required. Teachers in the treatment group will also be free to take part in other science staff development courses during the school year.

Control Group. Teachers who are randomly assigned to the control group will be informed that they have been placed on a waiting list for the course, and will be guaranteed placement in the course during the summer of 2009. The teachers in the control group will be free to take part in any science staff development that arises during the school year but will receive no treatment as part of this study. Assessment measures will be administered to control groups in each district at the same time that they are administered to the treatment group.

Data Collection Procedure and Timeline

Table A2 is a measurement schedule that describes each data collection activity, including its purpose, the sample involved, administration timeline and procedure. Each instrument (except for those copyright protected instruments[1]) is included in Appendix D1-D9) for review.

[1] ATLAST Test of Force and Motion for Students/Teachers – they are available for review upon request.

Table A2. Measurement Schedule, Samples, and Data Collection Procedures

Instrument	Purpose	Sample	Schedule	Procedure
Student measures:				
California *Standards Test* in Science	Science content knowledge (standardized test)	In 2008-09 school year, all physical science students in first two morning class sections of each participating teacher	Spring semesters (May 2008 & May 2009)	Administered by school; obtain scores from district.
ATLAST Test of Force and Motion for Students [Pre-test and post-test]	Knowledge of force and motion (project-administered)	All physical science students in first two morning class sections of each participating teacher	Before, and within two weeks after, force and motion unit (Nov 2008-Feb 2009)	Proctor administers in classroom.
Teacher measures:				
ATLAST Test of Force and Motion for Teachers [Pre-test and post-test]	Knowledge of force and motion	All participating teachers	Spring before study is run (May 2008) and spring one year later (May 2009)	District project coordinators administer to teachers in pre-study and end-of-year meetings.
Teacher Background Survey	Training, experience, and demographics; beliefs and practices related to teaching electric circuits before PD	All participating teachers	Spring semester (May 2008)	District site coordinators administer to teachers in pre-study meetings.
Teacher Post-Instruction Survey	Beliefs and practices related to teaching electric circuits after PD	All participating teachers	Spring semester (May 2009)	District site coordinators administer to teachers in end-of-year meetings.
Teacher Science-for-Teaching Interview	Pedagogical knowledge of force and motion for teaching	Random sample of 6 treatment and 6 control teachers' classrooms at 2 sites (n = 24)	Before the PD (summer/fall 2008) and after the force and motion unit (spring 2009)	Research staff interview teachers.
Classroom measures				
Classroom Observation Protocol	Classroom teaching practices and learning environment	Random sample of 6 treatment and 6 control teachers' classrooms at 2 sites (n = 24)	School year preceding (spring 2008) and following (spring 2009) PD course	Research staff observe and audio record classroom discourse during one class session pre and one class session post for each teacher.

Instrument	Purpose	Sample	Schedule	Procedure
Classroom Interview	Teacher pedagogical content knowledge and classroom teaching practices	Random sample of 6 treatment and 6 control teachers' classrooms at 2 sites (*n* = 24)	School year preceding (spring 2008) and following (spring 2009) PD course	In conjunction with classroom observation: research staff conduct interviews with teachers before and after classroom observations of one class session pre and one class session post for each teacher. The interview before the classroom observation is estimated to be 15 minutes long; the interview after the observation is estimated to take 30 minutes.

Science Content Tests.

Students' and teachers' science content knowledge will be measured using two sources of data: (a) scores on a state standardized test of students' science achievement, *California Standards Test (CST)*, and (b) force-and-motion tests developed and validated by the NSF-sponsored project, *Assessing Teacher Learning About Science Teaching* (ATLAST), by Horizon Research, Inc.

(a) State standardized test of students' science achievement. As with other state tests, all questions on the *California Standards Tests* are evaluated by committees of content experts, including teachers and administrators, to ensure their appropriateness for measuring the California academic content standards in middle school science. In addition to content, all items are reviewed and approved to ensure their adherence to the principles of fairness and to ensure no bias exists with respect to characteristics such as gender, ethnicity and language. Reported reliability figures for the CSTs in science range from .88-.91. Analyses will be performed both on total CST science scores, and on subscores reported on the Physical Science cluster and Force-and-Motion subcluster.

(b) Topic-specific student and teacher tests. To examine domain-specific effects of the teacher course, we will administer tests that were developed and validated as part of the ATLAST project in collaboration with Project 2061 of the American Association of the Advancement of Science (Smith & Banilower, 2006a, 2006b). ATLAST was funded by NSF to provide rigorous and well-validated measurement instruments to science education studies evaluating educational programs. In this study we will use ATLAST's Test of Force and Motion for Students and Test of Force and Motion for Teachers. These multiple-choice tests measure science content in the *National Science Education Standards*, and reflect the research literature documenting misconceptions related to science concepts in these domains.

Test of Force and Motion for Students. Based on national science content standards and the research literature on student and adult thinking and misconceptions about force and motion, the ATLAST project created a content framework that specifies learning goals, prerequisite knowledge, and misconceptions related to force and motion. The student test questions were specifically designed to assess conceptual knowledge in this framework, with no mathematics or other science concepts included. The developers intended the test to provide reliable information about students across a broad ability spectrum, and to be minimally burdensome. The test therefore is comprised of 25 multiple-choice items and is administered in one 45-minute period.

After defining the content domain, multiple-choice items were drafted that included distractors based upon research on student thinking and misconceptions. The test was validated to ensure that the items measure the intended assessment targets, beginning with cognitive interviews conducted with middle-school students. Items were revised based on these interviews, reviewed for content accuracy and domain coverage by physics experts, and further revised as needed. A pool of 60 items was then field-tested with over 5,000 students and 25 items were selected based on information provided by Item Response Theory (IRT) analyses. Dimensionality analyses revealed that items clustered in two sets; one representing general knowledge of force and motion, and one tapping a particularly prominent misconception—the idea that constant non-zero net force results in constant speed. IRT test information curves indicate that the final 25-item scale provides adequate or better information across a range of abilities from -3.0 *SD* to + 2.5 *SD* (Smith & Banilower, 2006a).

Test of Force and Motion for Teachers. The ATLAST content test for teachers assesses conceptual understanding in relation to the same conceptual framework as for students, related to force and motion, but in addition was constructed to assess domains of teacher knowledge for teaching this content. Although the 25 multiple-choice test items require science content knowledge, they are also all situated in instructional practice. Three types of items were included: knowledge of science content; using science content to analyze student thinking; and using science content to make instructional decisions. After extensive cognitive interviewing with teachers, 60 items were piloted in spring 2005. Dimensionality analyses revealed that the items clustered in one group, which was termed "content knowledge for teaching force and motion concepts." The analyses based on IRT were used to create a 25-item scale that spans the performance space and provides information across ability levels. The scale provides adequate information for ability estimates between about -2.5 *SD* and +2.5 *SD* (Smith & Banilower, 2006b).

Teacher Surveys

All teacher participants will be asked to complete a Teacher Background Survey in the spring preceding the intervention and a Teacher Post-Instruction Survey in the following spring after they have taught their classroom force and motion units.

The Teacher Background Survey contains questions about: (a) science and teaching education, teaching experience, and professional development experience; (b) when and for how long they taught force and motion during the previous academic year; (c) curriculum and textbook used to teach force and motion; and (d) demographics including sex, ethnicity, and race. Both the background survey and post-instruction survey include questions about factors that could mediate the relationship between knowledge gain from professional development experiences and impact on actual classroom practice, including teachers' reported classroom instructional practices when teaching science; attitudes about teaching science; and beliefs and attitudes about science, science teaching, and children's learning.

Teacher Science-for-Teaching Interview

Research staff will conduct in-depth interviews with teachers about their pedagogical beliefs and practices, with a specific focus on pedagogical content knowledge for science pedagogy specific to teaching force and motion. The interviewers will ask teachers open-ended questions about specific topics and problems in force and motion, including questions involving interpretation of student work. The post-unit interview also asks teachers to reflect on their professional development experience, and the extent to which it affected their teaching.

Classroom Observations

The purpose of the classroom observations is to gather descriptive, qualitative data on selected features of the classroom in order to understand the processes by which the teacher's professional development experience might influence student achievement. The aim here is not only to explore differences between groups, but also to inform the action theory of the professional development model.

We will focus on five areas of classroom teaching and interaction that are supported by the intervention and over which the teacher has influence, namely:

- Content accuracy;
- How teachers structure opportunities for talk about science and guide inquiry;
- How teachers encourage, respond to and incorporate student thinking, and anticipate student misunderstandings;
- How teachers foster student learning of science content, including their support for academic and English language development; and
- The variety and type of classroom interaction and participation structures.

The rationale behind focusing on these features is that they represent links in the chain of classroom influences that connect features of the Understanding Science professional development to student outcomes. The focus on exploration of science ideas in the courses builds teachers' own knowledge of the science content, which must result in science being taught accurately if students are to reach high standards. The focus on student thinking and conceptions in the courses stems from teachers' need for strategies to first access student ideas, and secondly build upon those scientific conceptions successfully during instruction. In the Understanding Science courses, participants read cases in which teachers assign tasks that reveal students' ideas about the science and facilitate sense-making conversation among students, thereby modeling for teachers ways of both eliciting student thinking and of providing opportunities for students to make sense of scientific ideas. Ultimately, the goal of collecting a set of descriptive exemplars is not only to illuminate the results of the study but also to provide models for teachers.

These classroom features will be described using a classroom observation protocol that has been adapted from (a) the 2002-03 Classroom Observation Protocol developed by Horizon Research, Inc., and (b) portions of the Center for Assessment and Evaluation of Student Learning (CAESL) Classroom Observation Protocol that focus on teachers' formative assessment behaviors, as well as (c) items designed to capture teacher focus on student thinking and aspects of student science talk. Initially we planned to use a combination of the Horizon and CAESL classroom observation protocols, but found that these alone would not produce the detailed qualitative observation data on classroom discourse and interaction necessary for this study.

The observation is intended to assess and describe how teachers' staff development affects their classroom practice, and how students develop academic language for science. For these reasons, the research team has developed an observation protocol that is designed to capture the variety of strategies and discourse practices that teachers may or may not develop as a result of difference professional development experiences.

Research staff who are trained in using the observation protocol will do the observations. Audio recordings will be collected during the observations for more detailed analyses of student-teacher discourse and student language use.

Classroom Teacher Interview

The classroom observations will be supplemented with teacher interviews about the specific force and motion lessons that are observed. These interviews will be conducted before and after each pair of observed lessons, will last about 15 minutes for pre-observation interview and about 30 minutes for post-observation interview, and will be conducted location and at a time that is convenient for the teachers but in the same week as the observed lessons.

The interviews include questions about learning goals and lesson planning for that specific lesson, the sequencing of activities, general questions about student ability level in that class, and what the teacher found particularly successful or challenging about the lesson. This information will help to orient the observer, and provide insights into how the teacher views the observed lesson compared to others in the force and motion unit.

A3. Use of Information Technology to Reduce Burden

Technology will be used in a variety of ways during the recruiting and data collection processes. During the recruiting period, brochures for teachers will list a web site at which teachers can obtain detailed information about the study at their convenience, and can apply to participate.

Basic contact information about the participants in the teacher courses will be collected online and stored in an electronic database created by the WestEd evaluation team. The evaluation team will use this database to keep track of course facilitator and teacher contact information and other information used to manage the study.

Second, communication between the evaluation team and selected school officials and/or teachers will occur through email, fax, and conference calls that take advantage of information technology and reduce burdens associated with paperwork. The communication will cover initial inquiries, the exchange of preliminary information, the scheduling and planning of site visits, and the review of draft reports.

Throughout the study, staff telephone numbers and email addresses will be available to respondents to allow them to contact the evaluation team with any questions or requests for assistance. This information, along with the names of contact persons on the evaluation team at HRA will be printed on all data collection instruments.

Student standardized tests scores from the California Standards Test will be collected through electronic databases at the district level.

A4. Efforts to Identify and Avoid Duplication

This is a new data collection, intended to provide scientific evidence as to the utility of a promising approach to teacher staff development. Although the model of staff development was designed to incorporate features of effective programs that are empirically associated with improvements in student achievement, the model is unique and data collected to evaluate other staff development programs do not inform the utility of the course being evaluated here. Thus, the proposed data collection does not duplicate any other data collections.

- The most similar information that is available comes from three years of evaluation studies of *Understanding Science* courses for elementary teachers. Conducted between 2000-2003, these studies identified positive teacher and student outcomes (Heller, Daehler, & Shinohara, 2003; Heller & Kaskowitz, 2004; Heller, 2006), and similar results have been found for middle-school courses in summer 2006. Funding limitations prevented implementation of the kind of randomized trials that would be necessary to establish causality between the teacher professional development courses and outcomes for participants and their students. While the evidence does not allow definitive conclusions because much of this work was not conducted at the grade level of the current proposal, and the samples were too small, the pattern of quantitative and qualitative findings suggests that gains were the result of teachers' participation in the *Understanding Science* courses.

In addition, each instrument used in the current study has been reviewed to make sure that we collect only the most necessary information to evaluate the outcomes of interest. The secondary information such as student standardized test scores will be accessed and collected through the electronic database at the school (or district) level.

A5. Impacts on Small Businesses and Other Small Entities

The evaluation team will collect data from few small entities, as most of the data sources will be from teachers and students. The few small entities are likely to be associated with the external technical assistants and consultants who may assist with data key-in and help with scoring the tests. Only minimal information will be needed from these small entities, and so no significant impact on small entities is expected.

A6. Consequences to Federal Programs or Policies if Data Collection is Not Conducted or Conducted Less Frequently

The data collection efforts in this study will allow researchers to study the impact of the *Understanding Science* model of staff development on student academic performance in science, as well as mediating teacher knowledge and skills for teaching science. Very few randomized control trials (RCTs) of this kind have been conducted to gather evidence about the effectiveness of teacher staff development programs, particularly in middle school science. A RCT is based on an experimental design, which is considered to be the strongest design when the interest of the study is in establishing a causal relationship (p.189, Trochim, W. M. K., 2001). The current study is an example of an RCT that investigates the relationship between a staff development program and various student outcomes. Failure to collect data using this kind of experimental design will greatly limit our capability of making cause-effect inferences about staff development implementation.

As depicted in Table A1, the research design relies on two administrations for each main data collection because the interest is in investigating whether the experimental treatment leads to post-treatment outcomes that are different from the control group, when controlling for initial differences between groups. If we collected data less often, such as post-treatment only, it would not be possible to make as rigorous inferences about group differences.

A7. Special Circumstances

This information collection fully complies with 5 CFR 1320.5(d)(2).

A8. Solicitations of Public Comments and Consultations with People Outside the Agency

a. ***Federal Register Announcement.*** The 60-day Federal Register notice was published on the following date: Vol. 72, September 4, 2007 page number 50675. To date there have been no public comments. A sample announcement is provided in Appendix E.

b. ***Consultations Outside the Agency.*** The evaluation team will seek the expertise of persons outside the agency through the creation of a Technical Working Group (TWG). The TWG will provide consultation on the design, implementation and analysis of this study, as well as the entire portfolio of Regional Educational Laboratory West (REL West) studies. They are expected to consult with REL West for five days per year through a combination of in-person and teleconferenced meetings. An honorarium of $1200 will be paid to each TWG member. This amount of honorarium is the daily rate required to retain TWG members. These are senior faculty members who are distinguished in their fields and extremely knowledgeable about the methodological and statistical requirements of randomized controlled trials. The rate and the contracts have been approved by IES.

   The TWG will play an important role in providing insight and guidance in support of a successful evaluation. The TWG members are listed below:

   - Professor Jamal Abedi, National Center for Research on Evaluation, Standards, and Student Testing CRESST, University of California, Davis
   - Dr. Lloyd Bond, Carnegie Foundation for the Advancement of Teaching
   - Professor Geoffrey Borman, University of Wisconsin
   - Professor Brian Flay, Oregon State University
   - Professor Tom Good, University of Arizona
   - Dr. Corinne Herlihy, Manpower Demonstration Research Corporation (MDRC)
   - Dr. Joan Herman, CRESST, University of California, Los Angeles
   - Professor Heather Hill, University of Michigan

- Dr. Roger Levine, American Institutes for Research (AIR)
- Professor Juliet P. Shaffer, University of California, Berkeley
- Dr. Jason Snipes, Council of the Great City Schools

A9. Respondent Payments

It is our experience that stipends are essential if we are to obtain the cooperation of participants with all data collection requirements, and to recruit and retain the number of teachers needed to conduct a study with sufficient power to detect policy-relevant impacts. Each teacher who completes both the 24-hour professional development course and 8-hours of program out-of-classroom activities will receive the base pay rate (approximately $25 per hour for outside of contract time) for participating in the professional development. Teachers will also be paid a stipend of $200 for the 4 hours of between-session assignments and 4 hours of data collection time outside of course sessions. Teachers who are randomly selected for, and who agree to participate in, the classroom observation and teacher interview studies will receive an additional stipend of up to $100.

REL West has six randomized controlled trials (RCTs) under way or in the OMB review process at this time. The incentive structures that have been approved include both those for data collection activities and the out-of-school (summer) hours that are required to participate in the studies. In several of the studies, compensation in the range of $25-$30/hour has been approved for data collection activities. As well, for the study most similar to Understanding Science, Study D: Problem-based Instruction in High School Economics, teachers (treatment and delayed treatment) participating in the 5-day professional development workshop did/will receive $1,000 (40 hours at $25/hour) for the work they contributed during summer out-of-contract activities.

A10. Confidentiality Assurances

HRA follows the confidentiality and data protection requirements of IES (The Education Sciences Reform Act of 2002, Title I, Part E, Section 183). HRA will protect the confidentiality of all information collected for the study and will use it for research purposes only. No information that identifies any study participant will be released. Information from participating institutions and respondents will be presented at aggregate levels in reports. Information on respondents will be linked to their institution but not to any individually identifiable information. No individually identifiable information will be maintained by the study team[2]. All institution-level identifiable information will be kept in secured locations and identifiers will be destroyed as soon as they are no longer required. HRA obtains signed NCEE Affidavits of Nondisclosure from all employees, subcontractors, and consultants that may have access to this data and submits them to our NCEE Contracting Officer's Representative (COR). Assurances of confidentiality are

[2] If any students are called by name on audio recordings, their names will be erased from the recordings within three days of the interview or class observation. No names will appear in transcripts of the audio recordings. Therefore, although it will be possible to hear students’ voices on the recordings, the voices will not be identifiable.

provided in consent forms for teachers, parents/guardians, and students (see Appendices F-H).

The following statements will appear verbatim on all letters, brochures, consents, and other study materials:

> According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid Office of Management and Budget (OMB) control number. The valid OMB control number for this information collection is xxxx-xxxx. The time required to complete this information collection is estimated to average xx minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Education, Washington, D.C. 20202-4651. If you have comments or concerns regarding the status of your individual submission of this form, write directly to: Rafael Valdivieso, U.S. Department of Education, 555 New Jersey Avenue, NW, Room 506E, Washington, D.C. 20208.
>
> Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

In all of our studies, we are committed to treating participants, be they teachers, students, or other educational personnel, as autonomous human beings, capable of making their own decision/choices. We are especially careful to protect children, who have limited autonomy because of age, by providing parents and guardians with decision-making prerogative. In accordance with this principle, we (a) always obtain informed consent in writing from all research participants and parents or guardians of minor student participants, and written assent to participate from students; and (b) protect the anonymity and confidentiality of all data and information collected from and about participants.

We also make every effort to minimize harms and maximize benefits for research participants and the broader educational community. In accordance with this principle, we (a) implement the best possible research designs to maximize benefits and minimize harms to teachers, students, and the field of education, and (b) only conduct studies that have a favorable risk-benefit relationship.

HRA will maintain data security protections for this study using procedures summarized in the following policies. Additional details of procedures are provided in Table A3.

Policies for Class 1 Data–Confidential data with identifying information

1. Can never leave HRA premises.
2. Always keep in a secure place.
3. Create separate working analysis files.
4. Only authorized persons can access and use.
5. Must be properly disposed of or transferred.

Policies for Class 2–Proprietary data and documents that are not Class 1

1. Must be used and stored under responsible person's oversight.
2. Must not be left in public view.
3. Create separate working analysis files.

4. Only authorized persons can access and use.
5. Must be properly disposed of or transferred.

Table A3. Procedures for Handling Class 1 Data

<table>
<tr><th>Policy</th><th>Electronic Data</th><th>Paper Data</th></tr>
<tr><td>Receive data promptly and track all materials.</td><td colspan="2">• Notify operations director and research support staff if expecting to receive confidential data.
• Catalogue all data received.</td></tr>
<tr><td>Can never leave HRA premises.</td><td colspan="2">• Must work on HRA premises with these data</td></tr>
<tr><td rowspan="2">Always keep in a secure place.</td><td>• May not be stored on laptops.</td><td></td></tr>
<tr><td colspan="2">• Store electronic files on data server and media in locked cabinet in locked room.
• Must not be left unattended in public view (e.g., on desk or screen).</td></tr>
<tr><td>Create separate working analysis files.</td><td>• Strip individual-identifying information from analysis files.
• Store in access-limited folders on HRA’s LAN.</td><td>• Remove cover sheets with respondent identifying information upon receipt by the research staff, leaving only ID numbers on the instrument.
• Store cover sheets in a secure location, separate from the completed instruments.</td></tr>
<tr><td>Only authorized persons can access and use.</td><td>• Limit access to the data server through use of passwords.
• Allow access by a minimum number of people who need to use the data.</td><td>• Key to locked cabinet to be kept securely by authorized persons.
• Allow access by a minimum number of people who need to use the data.</td></tr>
<tr><td rowspan="2">Must be properly disposed of or transferred.</td><td>• Mail data in a password protected and/or encrypted file on an unmarked CD or DVD.
•</td><td>• Shred any paper with confidential data before disposing.</td></tr>
<tr><td colspan="2">• Update catalogue whenever data are disposed of or transferred.
• When mailing, require recipient and delivery verification.</td></tr>
</table>

A11. Sensitive Questions

As shown in the information collection instruments in Appendix D, respondents will not be asked sensitive questions.

A12. Total Annual Cost Burden to Respondents

Table A4 lists the estimated time and cost burden associated with collection of information from respondents. Note that this is a three-year study. Therefore, the annual number of responses was obtained by dividing the total number of responses by 3 where the total number of responses equal to the number of respondents multiplied by the frequency of data collection. The annual unduplicated number of respondent is estimated to be 1,652[3].

[3] 1652 = (120 teachers + 6 district administrators + 25 school principals + 6 district data specialists + 4800 parents/guardians) / 3.

Table A4: Estimated Burden for Information Collection

Data Collection Activity	Number of Respondents	Frequency of Data Collection	Annual Number of Responses	Average Time Per Response (in minutes)	Annual Hour Burden	Est. Hourly Wage	Est. Annual Cost Burden
Teacher Measures							
Force and Motion Test for Teachers	120	2	80	40	53	$25	$1,325
Teacher Background Survey	120	1	40	30	20	$25	$500
Teacher Post-Instruction Survey	120	1	40	40	27	$25	$675
Teacher Science-for-Teaching Interview	24	2	16	60	16	$25	$400
Classroom Measures							
Classroom Observation Protocol	24	2	-	-	-	-	-
Classroom Interview	24	2	16	45	12	$25	$300
Total for Information Collection	-	-	192	-	128	-	$3,200

Table A5 lists the estimated time and cost burden associated with each task during the sampling and gaining cooperation process. The number of principals/teachers corresponds to the number of schools from which teachers will be recruited based on the study design and power estimates.

Table A5. Estimated Burden for Sampling and Gaining Cooperation

Task	Type of Respondent	Number of Respondents	Frequency of Activity	Annual Number of Responses	Time Estimate (in hours)	Annual Hour Burden	Est. Hourly Rate	Estimated Annual Cost Burden
Sampling tasks	District administrators	6	1	2	1	2	$45	$90
Gaining cooperation	School principals	25	1	8	1	8	$36	$288
Gaining cooperation	Teachers	120	1	40	1	40	$30	$1,200
Gaining cooperation	District data specialists	6	1	2	4	8	$30	$240
Obtaining consent from parents / guardians	Parents / guardians	4,800	1	1,600	0.2	320	$20	$6,400
Total for Sampling and Gaining Cooperation	-	-	-	1,652	-	378	-	$8,218

The estimated total annual number of responses and hour/cost burden for all data collection and recruitment activities is presented in Table A6.

Table A6. Annual Number of Responses and Hour/Cost Burden by Task and Total

Task	Annual Number of Responses	Annual Hour Burden	Estimated Annual Cost Burden
Information Collection	192	128	$3,200
Sampling and Gaining Cooperation	1,652	378	$8,218
Grand Total	**1,844**	**506**	**$11,418**

A13. Estimate of the Total Annual Cost Burden to Respondents or Record Keepers

There are no direct start-up costs to respondents other than their time to participate in the study, as estimated above. Estimations of the value of participation time for each respondent group, and for the study as a whole, are presented in Table A6 above.

A14. Annualized Cost to the Federal Government, Including a Description of the Method Used to Estimate Costs

The total cost for the study is $1,450,772 over three years. The average yearly cost is about $483,591. Most of the costs for the study are incurred in years 2008 and 2009 as data collection and processing efforts are under way.

A15. Program Changes or Adjustments

This program change is due to this being a new program.

A16. Tabulation, Analysis, and Publication of Results

A general timeline of the project is provided in Table A7 below. See Table A2 for detailed schedule for each data collection activity.

Table A7. Timeline for Evaluation Research

Sep 06 – Dec 07	**Jan 08 – Jun 08**	**July 08 – Aug 08**	**Sep 08 – May 09**	**Jun 09 – Aug 09**
Revise Study Design, Develop & Finalize Instruments, Revise Data Collection & Analysis Protocols, OMB & IRB Submission	School Recruitment, Informed Consent, Random Assignment, Baseline Teacher Data Collection	Professional Development for Treatment Teachers	Program Implementation, Process Data Collection, Post-intervention Data Collection	Post- intervention Data Collection (continued), Professional Development for Control Teachers, Data Analyses & Report

Quantitative Analyses

The analysis of treatment impacts will depend on the random assignment research design as its primary source of inference. Post-test outcomes for students and teachers in the treatment group will be compared to outcomes for students and teachers in the control group. The primary hypothesis-testing analyses will involve fitting conditional mixed-effects regression models (HLM or multilevel models), with additional terms to account for nesting of students within teachers (e.g., see Raudenbush & Bryk, 2002). Potential fixed effects include treatment group, baseline (pre-test) measures of outcome variables, other observed covariates, classrooms, and school site.

Multilevel regression models will be estimated to provide evidence related to the research questions. Specifically, students' science content knowledge, as measured by the two student assessments, will each be regressed on pre-test scores and demographic variables, and the experimental condition of the teacher. EL/English proficient differences in program impacts will be examined by including interactions of treatment status and EL status in the models. Analogous models will be estimated for teacher outcomes, except fixed effects for school will be included to account for the nesting of teachers within schools.

Consider the following hierarchical linear model for a continuous outcome:

$$Science_{ijk} = \alpha_0 + \beta_1 \mathrm{Pre}_{ijk} + \sum_{k=2}^{S} \beta_{2k} S_{jk} + \sum_{k=1}^{S} \beta_{3k} \left(S_{jk} Tx_{jk} \right) + \sum \beta_I I_{ijk} + \sum \beta_T T_{jk} + \tau_{jk} + \varepsilon_{ijk} \quad \textbf{[1]}$$

where subscripts *i*, *j*, and *k* denote student, teacher, and site, respectively; *Science* represents student science achievement; *Pre* represents the baseline measure of the outcome variable; *S* represents a dichotomous variable for each site, *Tx* is a dichotomous variable indicating student enrollment in a teacher's class who has been assigned to the treatment condition; and *I* and *T* are two vectors of control variables for students and teachers, respectively, measured prior to exposure to the intervention.[4] Lastly, $\boldsymbol{\tau}$ represents a variable for teachers (clustering group), and $\mathbf{e}_{ijk}$ is an error term for individual sample members. In this model, we estimate separate impacts for each site. The pooled impact estimate is calculated as the average of the site-specific impact estimates, and the standard error of the pooled impact estimate is calculated using information from the variance-covariance of the individual teacher impact estimates (see Dynarski et al. 2004). $\boldsymbol{\tau}_j$ captures effects (intercept) of teachers which accounts for the positive intraclass correlations in the data. This model allows us to assess the generalizability of the results across sites in the sample (Schochet 2005).

Note that model **[1]** is applicable to the analysis of the ATLAST Test of Force and Motion for students. Analyses of 2008/09 student standardized test results will rely on 2007/08 standardized test score data aggregated at the teacher level as a covariate. Thus, a teacher-level pretest will be included in the model examining standardized test scores.

4

Simple extensions to **[1]** allow us to examine the extent to which the intervention is associated with a reduction in the achievement gap between ELs and English proficient students or between low- and high achieving students (Research Question #2) by including interactions between treatment status and one of the variables in *I*. Model **[2]**, for example, which is a simplification of model [1] in that it does not estimate site-specific impacts, shows how we can estimate separate program effects for EL and non-EL students:

$$Science_{i:j} = a_0 + b_1 Pre_{ij} + b_{2E} Tx_j ELL_{ij} + b_{2N} Tx_j NonELL_{ij} + \sum b_I I_{ij} + \sum b_T T_j + \boldsymbol{\tau}_j + \mathbf{e}_{i:j} \qquad \textbf{[2]}$$

The only difference between this model and **[1]** is that the term $b_2 Tx_j$ is replaced by two terms that interact program variable Tx_j with dichotomous variables EL and non-EL students. Program impacts on EL and non-EL students are captured by the coefficients b_{2E} and b_{2N}, respectively. By statistically testing the hypothesis $b_{2E} = b_{2N}$, we can then establish whether program impacts reduce the achievement gap between EL and non-EL students.

Similar subgroup analyses will be possible across low and high performing students. To examine the extent to which the intervention reduces the achievement gap between low-performing and high performing students, we focus on student performance on the ATLAST force and motion achievement outcome. We use 7th grade ELA and Mathematics standardized test scores to classify students as low- and high-performing, with those at the “below basic” or lower level (≈ 35%) on either ELA or Mathematics classified as low performing, and those scoring above the “below basic” level classified as “high performing.” So classified, an analogous model to [2] will be estimated. We plan to examine the following student-level subgroup differences: 1) low- vs. high performance, 2) EL vs. English proficient, 3) racial/ethnic group, and 4) SES (free/reduced lunch).

Regression models will also be used to estimate program impacts on teachers’ content knowledge and instructional practices. For example, teacher post-test content knowledge will be modeled as a function of science pretest scores, treatment status, site-level fixed effects, and teacher covariates (e.g., years of teaching experience).

$$Science_{jk} = \alpha_0 + \beta_1 \mathrm{Pre}_{jk} + \sum_{k=2}^{s} \beta_{2k} S_{jk} + \beta_3 Tx_{jk} + \sum \beta_T T_{jk} + \varepsilon_{jk} \qquad \textbf{[3]}$$

Publications of Results

As indicated in Table A7, we plan to produce a technical/non-technical report in which evaluation results will be presented. In addition, the report will be published through the REL network and results of this research will be made publicly available through the Internet, on WestEd and HRA web sites, as well as through entries in the National Science Digital Library (NSDL) and Teacher Education Materials (TE-MAT) Project database. Finally, we anticipate making contributions to peer-reviewed journals and making presentations at research meetings.

A17. Approval Not to Display the Expiration Date for OMB Approval

No request is being made for exemption from displaying the expiration date.

A18. Exception to the Certification Statement

No exceptions to the certification statement are requested or required.

References

Birman, B., Desimone, L., Porter, A., & Garet, M. (2000). Designing professional development that works. *Educational Leadership, 57* (8), 28-33.

Bloom, H.S., Bos, J.M. and Lee, S.W. (1999). Using Cluster Random Assignment to Measure Program Impacts: Statistical Implications for the Evaluation of Education Programs. *Evaluation Review, 8* (23), 445-469.

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

Bodilly, S., Keltner, B., Purnell, S., Reichardt, R., & Schuyler, G. (1998). *Lessons from New American Schools' scale-up phase*. Santa Monica, CA: RAND.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn. Brain, mind, experience, and school: Expanded edition*. Washington, DC: National Academy Press.

Bravo, M., & Garcia, E. E. (2004, April). *Learning to write like scientists: English language learners' science inquiry & writing understandings in responsive learning contexts.* Paper presented at the annual meeting of the American Educational Researchers Association, San Diego, CA.

Brewer, D., & Goldhaber, D. D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22,* 129-145.

Broughman, S. P. & Rollefson, M. R. (2000). Teacher supply in the United States: Sources of newly hired teachers in public and private schools: 1987-88 to 1993-94. *Education Statistics Quarterly, 2* (3), 28-32.

Buxton, C. (1998). Improving the science education of English language learners: Capitalizing on educational reform. *Journal of Women and Minorities in Science and Engineering, 4* (4), 341–363.

Carlsen, W. S. (1991). Subject-matter knowledge and science teaching: A pragmatic perspective. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 2, pp. 115-124). Greenwich, CT: JAI Press.

Carlsen, W. S. (1993). Teacher knowledge and discourse control: Quantitative evidence from novice biology teachers' classrooms. *Journal of Research in Science Teaching, 30,* 471-481.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works.* New Haven: Yale University Press.

Collias, K., Pajak, E., & Rigden, D. (2000). *One cannot teach what one does not know: Training teachers in the United States who know their subjects and know how to* *teach their subjects. Washington, DC: Council for Basic Education.* Available at http://www.c-b-e.org/PDF/OneCannotTeach.pdf.

Dobb, F. (2004) *Essential Elements of Effective Science Instruction for English Learners,* 2nd edition. California Science Project: Los Angeles, CA.

Donner, A. N., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research.* London: Arnold.

Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching, 20* (5), 467-479.

Dynarski, M., Moore, M., Rosenberg, L., James-Burdumy, S., Deke, J., and Mansfield, W. (2004). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program, New Findings.* Final report submitted to the U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc.

Elmore, R. F. (1996). Getting to Scale with Good Educational Practice. *Harvard Educational Review, 66* (1), 1-26.

Fathman, A. K. & Crowther, D. T. (2006). *Science for English Language Learners: K-12 Classroom strategies.* Arlington, VA: NSTA press.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd (Ed.), *Holding schools accountable.* Washington, DC: The Brookings Institution.

Fulp, S. (2002). *Status of middle school science teaching: 2000 National survey of science and mathematics education.* Chapel Hill, NC: Horizon Research, Inc.

Gandara, Maxwell-Jolly, & Driscoll (2005). *Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs.* Santa Cruz, CA: The Center for the Future of Teaching and Learning.

Greenleaf, C. L.., Schoenback, R., Cziko, C., & Mueller, F. (2001). Apprenticing adolescent readers to academic literacy. *Harvard Educational Review, 71,* 79-127.

Hashweh, M. (1987). Effects of subject matter knowledge in the teaching of biology and physics. *Research and Teacher Education, 3,* 109-120.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 153-161

Hedges, L.V., & Hedberg, E.C. (2006) *Intraclass correlation values for planning group randomized trials in education*. IPR Working Paper, WP-06-12. Evanston, Ill.: Institute for Policy Research, Northwestern University.

Heller, J. I. (2006). *Final evaluation report for Science Cases for Teacher Learning: Impact on teachers, classrooms, and students, project years 2000-2003*. Submitted to WestEd and Stuart Foundation.

Heller, J. I., Daehler, K., & Shinohara, M. (2003). Connecting all the pieces: Using an evaluation mosaic to answer an impossible question. *Journal of Staff Development, 24,* 36-41.

Heller, J. I., Daehler, K. R., Shinohara, M., & Kaskowitz, S. R. (2004, April). *Fostering pedagogical content knowledge about electric circuits through case-based professional development.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Vancouver, B.C.

Heller, J. I., & Kaskowitz, S. R. (2004). *Final evaluation report for Science Cases for Teacher Learning: Impact on teachers, classrooms, and students, Project years 2000 – 2003.* Technical report submitted to WestEd and Stuart Foundation.

Hill, H.C. and Ball, D.L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education, 35* (5):330-351.

Lee, O. (2002). Science inquiry for elementary students from diverse backgrounds. In W. G. Secada (Ed.), *Review of Research in Education* (Vol. 26, pp. 23-69). Washington, DC: American Educational Research Association.

Lee, O., & Fradd, S. H. (2001). Instructional congruence to promote science learning and literacy development for linguistically diverse students. In D. R. Lavoie, & W-M. Roth (Eds.), *Models for science teacher preparation: Bridging the gap between research and practice* (pp.109-126). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J.R. Martin & R. Veel, (Eds.), *Reading Science* (pp.87-113). London: Routledge.

Lemke, J. L. (2002). Enseñar todos los lenguajes de la ciencia: palabras, símbolos, imágines, y acciones. In M. Benlloch (ed.), *La educación en ciencias.* Barcelona: Paidos. pp. 159-186.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers College Record, 105,* 913-945.

McLaughlin, M. (1991). The Rand Change Agent study: Ten years later. In A. Odden (Ed.), *Education policy implementation* (pp. 143-155). Albany, NY: State University of New York Press.

Merino, B.J., & Hammond, L. (2002). Writing to learn: Science in upper-elementary bilingual classroom. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 227-244). Mahwah, New Jersey: Erlbaum.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13,* 125-145.

Monk, D. H., & King, J. (1994). Multilevel teacher resource effects on pupil performance in secondary mathematics and science. In R. G. Ehrenberg (Ed.), *Choices and consequence.* Ithaca, NY: ILR Press.

Murray, D. M. (1998). *Design and analysis of group randomized trials.* New York: Oxford University Press.

National Center for Educational Statistics. (2002). *Schools and staffing survey, 1999-2000: Overview of the data for public, private, public charter, and Bureau of Indian Affairs elementary and secondary schools.* Available at http://nces.ed.gov/pubs2002/2002313.pdf

National Research Council. (2000). *Educating teachers of science, mathematics and technology: New practices for the new millennium* (Committee on Science and Mathematics Teacher Preparation). Washington, DC: National Academy Press.

National Assessment of Educational Progress: The Nation's Report Card (2006). *National and State Reports in Science.* Available at http://nationsreportcard.gov/science_2005.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, Sec. 901, 115 Stat. 1425 (2002). Retrieved at http://www.ed.gov/policy/elsec/leg/esea02/index.html.

Racine, D. (1998). Replicating programs in social markets. [On-line]. Retrieved at http://www.replication.org

Raudenbush, S. W. (1997). Statistical analysis and optimal design in cluster randomized trials. *Psychological Methods, 2,* 173-185.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Sage: Newbury Park, CA.

Redish, E. F. (1996). *Discipline-specific science education and educational research: The case of physics*. Paper prepared for the Committee on Developments in the Science of Learning for the Sciences of Science Learning: An Interdisciplinary Discussion, College Park, MD.

Rivers, J. C., & Sanders, W. L. (2002). In L. T. Izumi & M. E. Williamson (Eds.), *Teacher quality* (pp. 13-23). Stanford, CA: Hoover Institution Press.

Rosebery, A., B. Warren and F. Conant (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences, 2,* 61-94.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employee's performance to study the effects of teacher on students' achievement. *Sociology of Education, 70,* 256-284.

Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville: University of Tennessee.

Saxe, G., Gearhart, M., & Nasir, N. (2001). Enhancing students understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education, 4* (1), 55-79.

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

Schorr, L. (1997). *Common purpose: Strengthening families and neighborhoods to rebuild America*. New York, NY: Anchor Book Doubleday.Shulman, L. (1986).

Schweingruber, H.A. & Nease, A. A. (2000, April). *Teachers reasons for participating in professional development programs: Do they impact program outcomes?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Shinohara, M., Daehler, K. R., & Heller, J. I. (2004, April). *Using a pedagogical content framework to determine the content of case-based teacher professional development in science.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Vancouver, B.C., Canada.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15* (2), 4-14.

Shymansky, J. & Matthews, C. (1993). *Focus on children's ideas about science: An integrated program of instructional planning and teacher enhancement from the constructivist perspective.* The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Misconceptions Trust, Ithica, NY.

Smith, S. P. & Banilower, E. R. (2006a, April). *Measuring middle grades students' understanding of force and motion concepts: Insights into the structure of student ideas*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco.

Smith, S. P. & Banilower, E. R. (2006b, April). *Measuring teachers' knowledge for teaching force and motion concepts*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco.

Smith, L., Ross, S., McNeils, M., Squires, M., Wasson, R., Maxwell, S., Weddle, K., Nath, L., Grehan, A., & Buggey, T. (1998). The Memphis Restructuring Initiative: Analysis of activities and outcomes that affect implementation success. *Education and Urban Society, 30* (3), 296-325.

Thornton, R. K. & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics, 66*(4), 338-351.

Torres, H.N., & Zeidler, D.L. (2002) The effects of English language proficiency and scientific reasoning skills on the acquisition of science content knowledge by Hispanic English language learners and native English language speaking students. *Electronic Journal of Science Education, 6*.

Van Driel, J. H., Verloop, N., & De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 35* (6), 673-695.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Weiss, I. R., Arnold, E. E., Banilower, E. R., & Soar, E. H. (2001). *Local systemic change through teacher enhancement: Year six cross-site report.* Chapel Hill, NC: Horizon Research, Inc.

Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). *Report of the 2000 National Survey of Science and Mathematics Education.* Chapel Hill, NC: Horizon Research, Inc.

Weiss, I. R., Gellatly, G. B., Montgomery, D. L., Ridgway, C. J., Templeton, C. D., & Whittington, D. (1999). *Executive summary of the local systemic change through teacher enhancement year four cross-site report.* Chapel Hill, NC: Horizon Research, Inc.

WestEd & WGBH Boston. (2003) *Teachers as learners: A multimedia kit for professional development in science and mathematics.* Thousand Oaks, CA: Corwin Press.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11* (1), 57-67.

Appendices