J Sci Teacher Educ (2009) 20:415–436
DOI 10.1007/s10972-008-9120-9

Teaching for Understanding in Earth Science: Comparing Impacts on Planning and Instruction in Three Professional Development Designs for Middle School Science Teachers

William R. Penuel · Harold McWilliams · Carla McAuliffe · Ann E. Benbow · Colin Mably · Margaret M. Hayden

Published online: 2 December 2008

Abstract This paper compares and contrasts the impacts of three professional development designs aimed at middle school Earth science teachers on how teachers plan and enact instruction. The designs were similar in their alignment to research-based practices in science professional development: each design was of an extended duration and time span, included follow-up support to teachers, and incorporated active learning approaches in the professional development. In addition, the designs had a high level of coherence with other reform activities and with local standards. The main difference among the designs was in the roles of teachers in designing, adopting, or adapting curriculum materials. Evidence from teacher survey and observation data indicated that all programs had positive impacts on how teachers planned and enacted teaching for understanding, but differences among programs was more evident in their impacts on instructional planning.

Keywords Teaching for understanding · Professional development · Earth science · Middle school

W. R. Penuel (✉)
Center for Technology in Learning, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA
e-mail: william.penuel@sri.com

H. McWilliams · C. McAuliffe
TERC, Cambridge, USA

A. E. Benbow · C. Mably
American Geological Institute, Alexandria, USA

M. M. Hayden
Duval County Public Schools, Jacksonville, USA

Introduction

An enduring goal of science education of the last 50 years has been to develop student understanding of core scientific concepts by exposing them to well-designed curricular experiences (Atkin and Black 2003). Only a few years after the National Science Foundation first began investing in science curriculum development, Bruner (1960) argued that the goal of science education should be to give students "an understanding of the fundamental structure of whatever subjects we choose to teach" (p. 11). More recently, scholars have engaged in efforts to develop curriculum materials and other supports to help teachers *teach for understanding* (Cohen et al. 1993; Gardner and Dyson 1994; Treagust et al. 2001; Wiske 1997). The central premise behind this recent movement is that teachers should plan and enact instruction in which students have opportunities to learn about, experience, relate, and apply core disciplinary ideas (Gardner and Dyson 1994; Wiggins and McTighe 1998).

There is little doubt that professional development is necessary to prepare teachers to plan and enact instruction that develops students' deep understanding of subject matter. But beyond applying general principles from research (e.g., Loucks-Horsley et al. 1998) to designing effective professional development, few studies have explicitly compared different designs for achieving the aim of teaching for understanding. In fact, there are many possible designs for professional development programs that have been and can be created to prepare teachers to teach for understanding. Some important questions to answer about these programs are how do programs differ with respect to their designs and how do those differences matter, in terms of their effects on teachers' classroom practices?

This paper compares and contrasts three approaches to preparing teachers to teach for understanding in middle school Earth science with respect to both the *design* of the approaches and their *effects* on teachers' instructional planning and enactment in the context of an experimental study of teacher quality. All three designs reflected research-based principles for professional development, but they differed with respect to the role they gave to teachers in curriculum. In one design, teachers learned how to *adopt* high-quality curriculum materials developed by experts in Earth science and curriculum design. In a second design, teachers learned how to *design* curriculum experiences aligned to local standards using available materials and lessons they developed themselves. In a third design, teachers learned how to *adapt* expert-developed materials in a principled way to align to local standards. Survey and observational data provided evidence of the programs' impacts on instructional planning and enactment of a teaching for understanding approach in their classrooms.

Theoretical Framework

The growing body of empirical research on effective science professional development guided the theoretical framework for the study. Below, we review evidence for an emerging consensus about the importance of professional development that is of an extended duration and time span, includes follow-up for teachers, involves them in active learning, coheres with local standards and goals for student learning, and focuses on the content of instruction. We also point out that professional development models differ with respect to the roles teachers are expected to play in defining the content of instruction that is targeted by professional development. Here, there is less evidence to support a particular approach.

Duration and Time Span

A common criticism of professional development activities designed for teachers is that they are too short. Curricular reforms in science are extremely demanding and often require teachers to make big changes to implement them well (Bybee 1993; Crawford 2000). Frequently, the result is that teachers either assimilate new teaching strategies into their current repertoire with little substantive change or they reject those suggested changes altogether (Coburn 2004; Tyack and Cuban 1995). There is growing consensus that to make real changes, teachers need professional development that is interactive with their teaching practice, allowing for multiple cycles of presentation and assimilation of, and reflection on, knowledge (Anderson 2002; Blumenfeld et al. 1991; Jeanpierre et al. 2005; Kubitskey and Fishman 2006).

Professional development that is of longer duration and time span is more likely to contain the kinds of learning opportunities necessary for teachers to integrate new knowledge into practice (Brown 2004). For example, in their study of NSF-funded Local Systemic Initiatives, Supovitz and Turner (2000) found longer durations of professional development were needed to create “investigative cultures” in science classrooms, as opposed to small-scale changes in practice. Other large-scale studies of professional development have linked longer duration and time span to changes in teacher knowledge and practice (Desimone et al. 2002; Garet et al. 2001) and to higher levels of curriculum implementation (Penuel et al. 2007).

Follow-up

Even when they are of an extended duration, workshops and institutes rarely provide teachers with sufficient information and support for making changes to practice and for curriculum implementation. One reason why workshops are insufficient is that when teachers return to the classroom, they often encounter difficulties with planning and implementation that they have trouble solving on their own (Guskey 2002). Professional development staff associated with curricular innovations can support teachers through follow-up coaching and workshops to help them address their concerns (Penuel et al. 2005). Further, their efforts at follow-up are a means for monitoring implementation and for applying indirect pressure on teachers to try new practices associated with the professional development (Guskey 2002; Rowan and Miller 2007). When teachers experience this kind of follow-up, researchers have found teachers are more likely to make changes to their practice and to implement curriculum activities more consistently (Penuel and Means 2004; Penuel et al. 2007; Radford 1998; Tushnet et al. 2000).

Active Learning

Within science education, it is widely believed that to learn how to support student inquiry in the classroom, teachers need first-hand experiences of science in action either as part of their professional development or as part of apprenticeships to scientists (Gess-Newsome 1999). This need arises in part because most teachers today learned science from textbooks and tend to hold conceptions of the discipline and of how students learn that are inconsistent with how science knowledge actually unfolds through ongoing investigations by scientists (Boone and Kahle 1998; Marek and Methven 1991). Some research studies have presented evidence that supports the strategy of more hands-on, active learning, in that they have found a relationship between professional development activities in which teachers engage in inquiry and positive student achievement outcomes (Fishman et al. 2003; Jeanpierre et al. 2005).

There are other ways in which it may be important to promote teachers' active learning within professional development. Curriculum designers often have concerns about the ways teachers enact their curriculum, claiming some adaptations of materials constitute "lethal mutations" of those materials' intent (Spillane and Jennings 1997). It is this lack of understanding of underlying principles that some hypothesize prevents effective use of curriculum materials by teachers, especially those that rely on student-centered approaches to teaching (Lieberman and Miller 2001; Singer et al. 2000; Wiggins and McTighe 1998). The act of planning, enacting, and revising curricular units engages teachers more deeply with their teaching, so that they can come to understand more fully the principles of effective curriculum (Spillane 1999, 2004). It is not surprising, then, that research has found that professional development that incorporates time for instructional planning, discussion, and consideration of underlying principles of curriculum may be more effective in supporting implementation of innovations (Penuel and Means 2004).

Coherence

Coherence refers to teachers' interpretations of how well-aligned the professional development activities are with their own goals for learning and their goals for students. These interpretations are critical in at least two respects. First, teachers filter policy demands and messages from professional development about teaching through their own interpretive frames (Coburn 2001; Cuban 1986; Cuban et al. 2001). Second, the social context of schools has a strong influence on teachers' interpretive frames and thus their decisions about how to enact (or resist) particular innovations (Rivet 2006). If teachers perceive the demands to be aligned with their district's goals and with social pressures within the schools, they are more likely to perceive professional development focused on a particular innovation as congruent with their own goals, and thus commit to adopting or adapting the innovation (Lumpe et al. 2000). Similarly, teachers' perceptions about the level of administrative support for change are also likely to be related to their enactment of the innovation (Johnson 2007; Supovitz and Turner 2000). Past research has linked teachers' perceptions of coherence to changes in knowledge and practice (Garet et al. 2001) and to curriculum implementation (Penuel et al. 2007).

Content

There is widespread agreement that the content of professional development matters, and evidence from a wide range of studies supports this claim (Cohen and Hill 2001; Desimone et al. 2002; Garet et al. 2001; Hill et al. 2005; Penuel et al. 2007). Furthermore, when the content is closely linked to what teachers are expected to do in their classrooms, teachers are more likely to make use of what they learn, since it meets their needs for curricular activities they can use with students in the classrooms (Anderson 2002; Cohen and Hill 2001; Haney and Lumpe 1995; van Driel et al. 2001). When professional development content is also linked with specific curricular materials, those materials can be designed to extend what teachers are able to learn from formal professional development (Davis and Krajcik 2005; Schneider and Krajcik 2002).

Why Study the Design and Impact of Professional Development on Teachers' Practice?

The ultimate task of studies of the effectiveness of professional development is to demonstrate a link between professional development and student achievement (Loucks-Horsley and Matsumoto 1999). But the likely impact of any professional development on student learning is likely to be indirect: it will be a function both of the design of the program and of the design's impact on teachers' cognition and their instructional practice (Fishman et al. 2003). In this model of impact, it is first necessary to establish important ways that designs aimed at the same broad goal can vary and that the designs can be enacted as intended. The next step would be to establish whether the designs as enacted have their intended impacts on teacher cognition and on teacher practice in the classrooms. As we argue below, none of these can be assumed ahead of time but have to be validated in the field, before it is wise to invest in costly studies of impact that require large numbers of participants to achieve adequate power.

Consider first how designs might vary with respect to professional development aimed at supporting the adoption of particular curriculum materials. Beyond this consensus regarding the importance of content, professional development models vary widely with respect to the role teachers are expected to play in shaping the content of the teaching that is the focus of professional development. Traditionally, models of professional development have focused on preparing teachers to implement specific curricular materials, without adaptation. More recently, however, a number of projects have explored how curriculum revision or planned curriculum adaptation may be used to promote the improvement of teaching quality, to enable high-quality implementation, and to increase student achievement (e.g., Linn et al. 1993; Singer et al. 2000). Still other models of professional development put teachers in the role of designers of curriculum and professional development experiences (e.g., Lotter et al. 2006; Wiggins and McTighe 1998). Evidence from large scale correlational studies suggests that the differences in design of professional development with respect to how teachers' roles are conceived make a difference in terms of their effects on teacher knowledge and practice more broadly (Garet et al. 2001) and on the implementation of specific curriculum (Penuel et al. 2007).

Just as designs vary, so, too, can their effects on teacher cognition and classroom practice? Evaluators of program have long observed that it is necessary to study enactment, since the scale, depth, and fidelity of implementation cannot be assumed ahead of time (Patton 1979; Rossi et al. 2004; Scheirer 1994). Further, knowledge of how enactment varies is critical to understanding how variations help explain differences in effectiveness within and between programs (Lipsey and Cordray 2000). Studies of schools implementing different whole school reform designs found that different designs can and do differ in how easily that they are enacted by school leaders and professional developers (Bodilly 1998). These differences are linked in part to how well-specified the designs are, in terms of their intended effects: the "classroom footprint" of designs that provide less instructional guidance to teachers about what is expected of them is harder to discern from records of practice than for designs that provide clearer guidance to teachers (Correnti and Rowan 2007). Ultimately, differences in enactment among programs may also be linked to differences in effectiveness in improving student achievement (Supovitz and May 2004).

To date, studies of professional development have not compared designs in the way that researchers studying whole school reform designs have with respect to their differences and effects of those differences on teachers' practice. Instead, researchers studying the design of professional development and its impacts have tended to adopt a single approach for their project and study its effects on teaching and learning. This approach precludes comparing whether one design is more easily adopted to achieve a particular goal and whether one design is better at changing teacher knowledge than another.

In the study on which we report on this paper, professional developers created three different designs for a single large urban district, all aimed at meeting the district's goal to prepare teachers to teach for deep understanding. As the first part of an experimental study that is testing the impacts of those designs on teaching and learning, the lead author of this article worked with staff developers to specify similarities and differences among the designs. In the next section, we discuss these professional development designs using the theoretical framework detailed above. Then, in the method and results section, we analyze whether the differences and similarities among the designs could be detected in teachers' reports of the impact of the professional development on their cognition as reflected in their instructional planning process and in observations of their classroom practice.

The Professional Development Designs

Below, we describe each of the three designs with respect to its expected time span and duration, nature of follow-up, use of active learning strategies, coherence with the district standards, and content.

Preparing Teachers to Adopt the Investigating Earth Sciences Curriculum

Investigating Earth Systems (*IES*) is a 10-module middle school curriculum, funded by the National Science Foundation and developed by the American Geological Institute (AGI). The inquiry-based Earth systems science curriculum consists of a student edition with investigations and content; a teacher's edition with science background, students' misconceptions, teaching tips, materials management advice, assessments, *National Science Education Standards*-alignment; and online teaching resources. *IES* was written by a team of curriculum developers, scientists and teachers and was pilot- and field-tested over 3 years by middle school teachers across the United States. It was first published by It's About Time/Herff Jones Publishing in 2001, and has been adopted by the state of California, as well as such major school districts as Denver Public Schools, Chicago Public Schools, and the Clark County School District in Nevada (Las Vegas).

The content of the modules is organized around five "Big Ideas" in Earth science, but it is important to note that professional development for *IES* has a specific purpose: it is designed to prepare teachers to implement specific modules that fit a school district's middle school Earth science standards. In this way, through the selection of particular modules, *IES* attempts to provide materials that will allow teachers to meet their state's or district's requirements with respect to content coverage. In DCPS, AGI worked with district leaders to select the four modules that were most closely aligned to the Sunshine State Standards: *Dynamic Planet* (sixth grade), *Rocks and Landforms* (sixth and seventh grades), *Water as a Resource* (seventh grade) and *Astronomy* (eighth grade).

For the current study, AGI staff provided a 2-week initial workshop to all teachers assigned to the *IES* condition. The first part of the training covered topics that underpin the curriculum: typical module structure, nature of inquiry-based science and the Earth systems approach, managing materials and students working in collaborative groups, teacher support, *IES* website and assessment components used in *IES*. In the second part of the summer training, teachers worked in specialist groups to focus on activities and content from the particular *IES* modules they would be using with their students. During four follow-up training sessions throughout the academic year, AGI staff met with teachers to discuss issues and successes they experienced during the implementation. Teachers also had the opportunity to share student work and assessments and discuss adaptations they made to accommodate their students' ability levels.

Active learning strategies are evident in *IES* workshops in that leaders introduce teachers to inquiry principles early in the training and then give teachers frequent opportunities to set up and try hands-on investigations, both in the roles of students and as teachers. The training also includes practice with managing materials, setting up and using student journals as assessment tools, and using investigations as performance assessments. The teacher's edition of *IES* plays an important role in training, as teachers are encouraged to use the components of the teacher's edition to plan, implement, assess, and reflect upon their Earth science instruction.

With respect to coherence, both the fixed content of the modules and practical realities of implementation limit the degree to which perfect alignment to local goals could be achieved through the professional development. Significantly, teachers in sixth and seventh grade each taught a "partial" module (*Rocks and Landforms*) since the module targeted concepts that spanned two grade levels of standards. The alignment to Duval's local standards was, moreover, imperfect, in that some concepts (e.g., long-shore currents) were not in the modules themselves. For these topics, AGI staff identified web-based resources to supplement module content. In addition, AGI staff note that during the first year of implementation, teachers tend to follow the modules strictly, without adding or supplementing content. Therefore, they could be expected to make limited use of these supplemental resources, as they struggle with new, student-centered methods of instruction.

Earth Science by Design

Earth Science by Design (*ESBD*) is a year-long program of professional development created by TERC and AGI with funding from the National Science Foundation. *ESBD* prepares teachers to apply the principles of *Understanding by Design* (UbD) (Wiggins and McTighe 2005) to the teaching of Earth system science. The primary goals of the *ESBD* program are (a) to teach for deeper understanding by focusing on "Big ideas" and using an "Earth as a system" approach, (b) to design and apply appropriate assessment techniques, such as preconception surveys and authentic performance measures, and (c) to use visualizations and satellite imagery to promote student understanding. Teachers completing the *ESBD* program reorganize existing curricular materials, such as those from their textbooks or those they may have developed themselves or collected from colleagues at professional conferences, into coherent units of instruction that target essential questions and enduring understandings and that culminate with a performance assessment. Each teacher participating in the *ESBD* program is expected to reorganize one unit of instruction that they implement with their students. During this study, sixth grade teachers collaborated to organize a 9-week unit focused on the dynamic nature of planet earth (i.e., earthquakes, volcanoes, plate tectonics, forces of change). Seventh grade teachers organized a unit around water and eighth grade teachers had astronomy as their unit topic.

In addition to an initial 2-week workshop, teachers in the *ESBD* condition participated in 2 days of follow-up professional development in the fall after the summer workshop and 3 days of follow-up in the spring. The model calls for local staff developers to lead these activities; consistent with that model, staff from the district led the workshop and fall follow up activities, without any assistance from developers of the program. The 3 days in the spring included 1 day in which staff mentored teachers as they revised and edited their final unit plans and reflected on their implementation results. The other 2 days comprised a conference in which teachers gave presentations about their units. In addition, teachers received mentoring from staff during the school year, which consisted of help with the design of their units and help with managing other aspects of their participation in the study.

The content of the *ESBD* condition focused on the UbD approach to curriculum development. During the first week of the *ESBD* summer workshop, teachers engaged in activities and discussions to consider the nature of understanding, to struggle with what is worthy of understanding, and to begin to understand the "Earth as a system" approach to Earth system science. They also learned the process of "backward design" and practiced constructing a unit using the *ESBD* online unit planner. *ESBD* teachers gained practice with developing assessments of student learning intended to "convict" students of understanding.

During the second week of the workshop, *ESBD* teachers were given time to work collaboratively to produce the unit that they would be implementing the following school year. Teachers began by drafting essential questions and enduring understandings that their units would target. Next they developed a performance assessment that would reveal students understandings (as well as misunderstandings) of the unit. Last, by considering the lessons they had used when they taught the unit in previous years, they began to reorganize their units. *ESBD* teachers included activities and laboratory exercises in their units, only if the content of the activity or exercise directly targeted essential questions and enduring understandings. Often, teachers removed a lot of unnecessary lessons from their units and had to supplement them with lessons from professional resources other than their textbooks. Teachers were required to incorporate visualizations and Internet resources into their units.

Opportunities for hands-on practice in the *ESBD* condition's professional development activities took the form of trying out visualizations and engaging in unit planning. Teachers began the second week of the summer workshop reviewing an *ESBD* unit and engaging in one of the unit's activities that provided experience working with satellite data. Throughout the second week, teachers had time to craft their units, with mentoring from one of the two district staff. They also spent significant amounts of time trying out visualizations and Internet resources. In addition, they received feedback on their unit plans from colleagues in the workshop. On days 6, 7 and 8, three video presentations from previous *ESBD* teachers were used to engage teachers in whole-group discussions focused on the challenges of implementation.

The *ESBD* condition had the greatest potential coherence of professional development with teachers' local context, since teachers began with the standards and could use any materials whatsoever in their units that were aligned to those standards. District staff helped teachers map the Florida Sunshine State Standards and the district standards to the Enduring Understandings and Essential Questions for their units. At every possible opportunity during the professional development program, staff emphasized that teachers should make sure that their goals for students were aligned with the standards. In addition, teachers worked in small groups on their units, collaborating with other teachers who had responsibility for teaching the same standards that they did.

The *Hybrid* Approach

Teachers in the *Hybrid* condition participated in a year-long program of professional development comprised of a 2-week summer workshop, 2 days of follow-up professional development in the fall after the summer workshop, and 3 days of follow-up in the spring. AGI and TERC staff collaboratively led the workshop and fall follow up activities. In spring, the 3 days were led by DCPS staff and included 2 days of a spring conference in which teachers gave presentations about their units. In addition, teachers received mentoring from district staff during the school year, which consisted of help obtaining teaching materials and kits, help with the design of their units, and help with managing other aspects of their participation in the study.

The content of the *Hybrid* condition professional development blended content of the *IES* workshop and the *ESBD* workshop. Like the teachers in the *ESBD* workshop, teachers in the *Hybrid* condition engaged in activities and discussions to consider the nature of understanding, to struggle with what is worthy of understanding, and to begin to understand the "Earth as a system" approach to Earth system science. They also learned the process of "backward design" and practiced constructing a unit, just as the *ESBD* teachers did, using the *ESBD* online unit planner. Like *ESBD* teachers, *Hybrid* teachers also gained practice with developing assessments of student learning intended to "convict" students of understanding. But unlike the *ESBD* teachers, the *Hybrid* teachers made use of the particular *IES* modules that were aligned to their grade level in constructing their units. Moreover, teachers received instructions that at least 50% of the *IES* investigations were to be used in constructing their units.

Throughout, UbD concepts underlying the design of the *IES* materials were emphasized. For example, on day 3 of the workshop, AGI and TERC staff introduced the idea of "essential questions" (part of the UbD framework and the *ESBD* Summer Institute Guide). Teachers worked in groups of four to brainstorm essential questions. In addition to creating their own Essential Questions, teachers recorded the "key questions" from the *IES Rocks and Landforms* module into their brainstorming work. After reviewing the candidate Essential Questions, each group selected four to incorporate into their sample unit.

Opportunities for hands-on practice in the *Hybrid* condition's professional development activities took the form of practicing *IES* investigations and engaging in unit planning. During the second week of the summer workshop, teachers had time to craft their units, with mentoring from one of the three facilitator leaders. In general, they worked on their units in the mornings and in the afternoons engaged in hands-on investigations from the *IES* modules in order to familiarize themselves with these activities. They were able to ask questions of the *IES* facilitator and to become familiar with activities that they might wish to incorporate into their units. Approximately 45 min were set aside each afternoon for whole-group discussion of progress, problems, ideas, and issues that were emerging.

To increase the coherence of professional development with teachers' local context, staff helped teachers map the Florida Sunshine State Standards to the Enduring Understandings and Essential Questions for their units. In this activity, professional development staff emphasized that teachers should not start with the standards but rather make sure that their goals for students were aligned with the standards. In addition, teachers worked in small groups on their units, collaborating with other teachers who had responsibility for teaching the same standards that they did.

The Current Study

The current study compared the impacts of the three designs described above on how teachers plan and enact their Earth science units. It is part of a larger study that is comparing the impacts of these different designs on teachers, teaching practice, and student learning. Participants completed surveys that asked teachers to reflect on how the designs affected their instructional planning process after they had completed their units. While teaching their units, researchers conducted observations of teachers' practice, with the aim of documenting the alignment of teachers' practice to the goal of teaching for understanding. Descriptive and inferential statistics provided the research team with data to use to help interpret subsequent study results and the professional development team with data for judging the degree to which the enacted workshops were consistent with the professional development designs.

In this study, we asked:

1. What impacts did each design have on teachers' instructional planning for their Earth science unit, and how did those impacts differ by professional development program?
2. What impacts did each design have on teachers' classroom enactment of a teaching for understanding approach in their units, and how did those impacts differ by professional development program?

Using the professional development designs and correlates of effective professional development as guides, we developed the following hypotheses about the impacts of the designs on instructional planning and enactment:

H_1: More teachers in the *ESBD* and *Hybrid* conditions will report changes to their instructional planning, when compared with the *IES* condition, since *ESBD* and *Hybrid* designs both give strong emphasis to the process of instructional planning.

H_2: Relative to teachers in the control condition, teachers in all conditions will be more likely to use instructional strategies that are associated with a teaching for understanding approach, since teachers in all three conditions will experience professional development of a significant duration that is focused on implementing a UbD approach to teaching.

H_3: There will be no observed differences among conditions with respect to observed differences in enactment of instructional strategies linked to the UbD approach, since each design shares that approach to teaching for understanding.

Research Methods

The overall study used an experimental design, in which teachers were randomly assigned either to one of the three (professional development) treatment conditions or to the control condition. Random assignment studies have the fewest threats to internal validity, and are thus more likely to yield unbiased estimates of potential impact compared with other designs (Shadish et al. 2002). The random assignment process took place after teachers volunteered to be in the study; therefore, it is important to note that the findings of this particular study can be generalized only to groups of teachers who volunteer for professional development. Other efficacy studies that study the impact of the interventions when teachers were compelled to participate would be needed to establish the potential under those conditions.

District Context

All research participants were teachers in the Duval County Public Schools (DCPS). DCPS currently serves 125,820 students in 164 schools, of which 28 were middle school, all of whose science teachers were eligible for participation in the study. Sixteen of the 28 middle schools have a 50% or higher rate of eligibility for free or reduced-price lunch.

The district science curriculum for middle school adheres to the Florida Sunshine State Standards, which the State of Florida mandates all teachers must follow. The district has organized the standards into 9-week units, and Earth science standards are taught as part of one unit at each of the three middle school grades. Each grade level has one or more units based on earth space science concepts taught within a 9-week span. The district has "translated" the standards into Enduring Understandings (following the UbD model) and linked those understandings to each unit. Testing in science takes place only in eighth grade, and Earth and space science concepts cover roughly 25% of the items on that test.

Research Participants

A total of 41 sixth, seventh, and eighth grade teachers from 19 middle schools in a large urban district were assigned to one of the three professional development conditions. Teachers who volunteered represented three magnet schools (arts, science and math, and academic) and 10 schools with over 50% of the students on free or reduced-price lunch. Three middle schools that had a science teacher leader on staff had five or more teachers volunteer for the study.

Of the teachers that volunteered, 14 teachers were assigned to the *IES* condition, 13 to the *ESBD* condition, and 14 to the *Hybrid* condition. The differences among groups on the characteristics presented below in Table 1 are not statistically significant as measured by chi-square tests. Although some differences appear quite large in the table, the small sample size makes it difficult to achieve statistical significance for all but the largest of differences.

Sources of Data

Unit Implementation Questionnaire

We collected data on teachers' reported changes to instructional planning through an online questionnaire completed 1 year after their workshop. The questionnaire focused on a range of topics related to teachers' instructional planning process and unit implementation. The analysis presented in this paper focuses on qualitative responses to an item asking them to describe how the professional development influenced their instructional planning process. Although this approach is limited by

Table 1 Characteristics of faculty respondents to questionnaire

	Condition		
	IES	*ESBD*	*Hybrid*
Gender (%)			
Male	25.0	45.5	27.3
Female	75.0	54.5	72.7
Race/Ethnicity[a] (%)			
White	75.0	45.5	54.5
African American	16.7	45.5	36.4
Hispanic/Latino	8.3	0.0	18.2
Asian	0.0	0.0	0.0
Other/Unknown	9.0	0.0	0.0
Teaching experience (years)			
Years teaching	$M = 12.7$	$M = 14.4$	$M = 5.85$
	SD = 11.0	SD = 11.8	SD = 4.1
Years teaching science	$M = 10.5$	$M = 9.1$	$M = 4.3$
	SD = 8.2	SD = 5.9	SD = 3.1
Highest degree (%)			
Bachelor's	66.7	81.8	90.9
Master's	16.7	18.2	9.1
Educational specialist's	8.3	0.0	0.0
Missing	8.3	0.0	0.0
Teaching assignment			
6	5	5	5
7	3	3	4
8	4	3	2

[a] Teachers could select multiple categories

the fact that teachers' retrospective accounts may be biased or inaccurate, we know of no other easily implemented method for analyzing teachers' actual planning process other than soliciting their own thinking about their process.

Structured Observation Protocol

We constructed an observation protocol designed to measure the extent to which teachers' instruction was aimed at teaching for deep understanding in science. Whether or not teachers had students engage in activities to develop three of Wiggins and McTighe's (1998) *facets of understanding* was one focus of the protocol: observers recorded whether as part of instruction students generated explanations of scientific phenomena (explanation facet), interpreted, used, or judged models or analogies (interpretation facet), or applied apply concepts to solve a problem or to construct a product (application facet). In addition, observers rated the extent to which students interviewed as part of observations understood of the purpose of the day's instructional activities and the extent to which teachers elicited and made use of student preconceptions or ideas about the topic in their lessons. These ratings were on a scale from 0 ("not at all true of this class") to 3 ("very true of this class"). All members of the team, including researchers, professional developers, and district staff agreed that of these different aspects of instruction, the most important was whether students understood the purpose of the instructional activity: at the heart of the model of instruction each intervention is the idea that students are not simply passive recipients of instruction but are actively engaged in making sense of what they are doing. In that respect, students' ability to articulate to researchers why the teacher has assigned a particular task is critical to teaching for understanding.

As part of the study, a researcher visited each classroom once during the implementation of teachers' Earth science unit. Visits took place in the middle of units, within a 1-week period (all teachers were in roughly the same place in their units, ensuring comparability of observations across classrooms). All observers were trained ahead of time in the protocol and had practiced using the protocol in the field with classrooms not in the study. In addition, while in the field, observers conducted a total of 15 observations as pairs, allowing for the calculation of the percent of inter-rater agreement. With respect to facets of understanding, inter-rater agreement was between 80 and 100%. Ratings of the two observers with respect to their judgments of students' understanding of the purpose of the instructional activity were exactly the same in 66.7% of the observations and within one point in 100% of the observations. Ratings with respect to judgments about the extent to which teachers elicited student preconceptions were the same in 86.7% of observation and within one point in all observations.

Although teachers were randomly assigned to condition, there were multiple teachers within schools, so we used hierarchical linear modelings (Raudenbush and Bryk 2002) to test the impact of assigned condition on teachers' instructional practice. We modeled treatment condition as a random effect that varied among schools (Level 2 in our models); our interest, however, was interpreting the significance of teachers' treatment assignment on the different aspects of instruction (Level 1 in our models). Because the dependent variables included a mix of dichotomous and continuous variables, we used two kinds of models. For dichotomous variables, outcomes were modeled as log-odds and predictors as dummy variables representing the different treatment conditions (the control group was the omitted condition). The basic model for dichotomous variables testing the differences between the treatment and control conditions, shown below for the log-odds of observing a teacher calling for students to explain their ideas is:

Level 1 Model

$$\mathrm{Prob}(\mathrm{EN\_FACEX}_{ij} = 1/\beta_j) = \varphi_{ij}$$

$$\mathrm{Log}[\varphi_{ij}/(1-\varphi_{ij})] = \eta_{ij}$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}\left(\mathrm{IES}_{ij}\right) + \beta_{2j}\left(\mathrm{ESBD}_{ij}\right) + \beta_{3j}\left(\mathrm{HYBRID}_{ij}\right)$$

Level 2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

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

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

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

where η_{ij} is the log-odds of observing a teacher calling on students to explain their ideas; β_{0j} is the log-odds of observing a control teacher in school j calling for explanation; β_{1j} is the log-odds of observing an *Investigating Earth Systems* teacher in school j calling for explanation; β_{2j} is the log-odds of observing an *Earth Science by Design* teacher in school j calling for explanation; β_{3j} is the log-odds of observing a *Hybrid* teacher in school j calling for explanation; γ_{00} is the average log-odds in control schools for this outcome; γ_{10} is the average log-odds in *Investigating Earth Systems* schools for this outcome; γ_{20} is the average log-odds in *Earth Science by Design* schools for this outcome; γ_{30} is the average log-odds in *Hybrid* schools for this outcome; $u_{.j}$ difference between the observed log-odds and school averages.

The basic model for continuous variables differ from the model shown above in that the outcome is modeled using the original scale, and there is an error term, r_{ij}, associated with each individual teacher (in addition to the error term for schools). To compare treatments to each other, we computed separate models for which a different treatment condition was the excluded dummy variable each time. In the results section, we display all results in their original metric; however, significance levels reported reflect the results of the HLM models, since those properly estimate the effects of clustering of teachers within schools.

Results

Influence on Instructional Planning

As promoted in the professional development models, teachers in both the *ESBD* and *Hybrid* conditions reported that their participation in the project had had a large impact on their instructional planning process. Six of the *ESBD* teachers and six of the *Hybrid* teachers mentioned thinking more about what their students should know at the end of the unit when planning their units. In other words, these teachers gave much more weight to the "enduring understandings" they expected students to have at the end of their units:

> The experience has taught me to "look down the road" first, to determine where I want the students to end and then to determine how they are going to get there, starting with the "end" in mind. (ESBD Teacher)

> When planning instruction I'm looking at a large unit with common goals and all activities aligned, instead of a piecemeal week-to-week fashion. (IES Teacher)

A few teachers in both these conditions also mentioned that they made use of the data from preconceptions assessments given at the beginning of units to adjust their planned sequence of activities:

> I review students' preconception responses to determine what I'd emphasize. TIDES has given me a deeper understanding of how to engage my student incorporating technology and visualizations. Its also help me to effectively uncover deep understanding, misunderstandings, and preconceptions my students developed. (ESBD Teacher)

Finally, teachers in these two conditions also reported that they learned from the TIDES workshops *how* to go about planning.

> I have a much better idea of how to go about planning. At first, I was not sure if how I was doing the thing was right, wrong, or if anyone new exactly how to plan. However, I am now more confident in how I do my planning and rely much more on my own material. (Hybrid Teacher)

> I planned the entire unit before actually teaching this year, rather than planning as I go as I had done in previous years. (Hybrid Teacher)

> My planning process became more of a "reverse sequence" method. (ESBD Teacher)

Teachers in the *IES* condition did report that participating in the project had caused them to plan to incorporate more student-centered and hands-on laboratories and investigations in their teaching.

> I am leaning more towards "inquiry-based" science. Students need to be more responsible for discovering and teachers need to stop hand feeding students information. (IES Teacher)

> I am using more hands-on activities to keep students interested and focused. (IES Teacher)

These self-reported changes to instructional planning are consistent with the models of professional development they seek to promote. Both the *ESBD* and *Hybrid* workshops introduced teachers to a new approach to instructional planning. Although the two conditions differed in that the *Hybrid* teachers received curriculum materials with extensive opportunities for student investigations and the *ESBD* teachers did not receive these materials, teachers' reported changes to their instructional planning process were remarkably similar across conditions. The teachers in the *IES* condition had a distinctive profile, reflecting their efforts to incorporate more inquiry-oriented, hands-on activities in science with their students.

Notably absent from the *ESBD* and *Hybrid* teachers' reported changes to instructional planning were descriptions of new culminating performance tasks. Although the second step in the planning process emphasized in the workshops involves the design of a measure to assess enduring understandings that are the focus of the units, these did not figure in teachers' reported changes to practice. This fact suggests either that these aspects of the model may have been less salient for teachers or that teachers in the study had not adopted these model aspects.

Impacts on Enactment of Instruction

With respect to the probability of observing a class in which teachers engaged students with developing different facets of understanding, the designs were most effective in promoting the aspect of interpretation (Fig. 1). Students of teachers who participated in both the *ESBD* and *Hybrid* programs were significantly more likely to be observed interpreting, using, or judging models or analogies in science class than were students in control classrooms. There were no statistically significant differences between the treatment and control conditions with respect to the probability of observing students giving explanations or applying something they had learned, and there were no statistically significant differences among the program designs on any of the measures of facets of understanding.

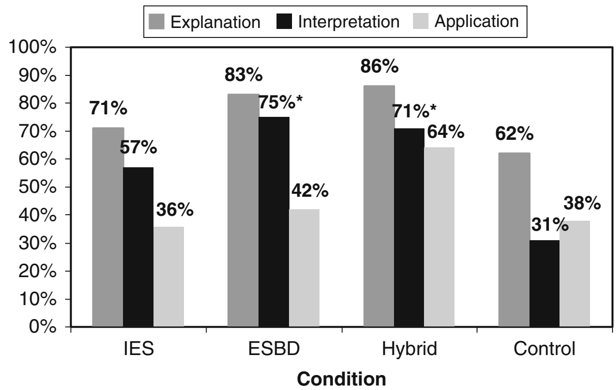

Fig. 1 Student engagement with facets of understanding. * Significantly different from the control condition at $p < 0.05$

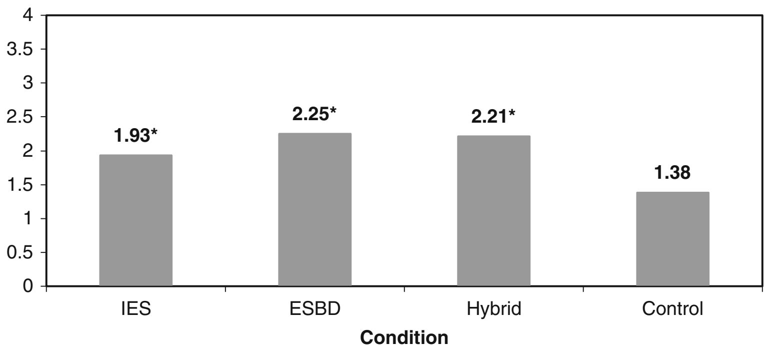

Fig. 2 Observer ratings of how well students understood the purpose of the instructional activity. * Significantly different from the control condition at $p < 0.05$

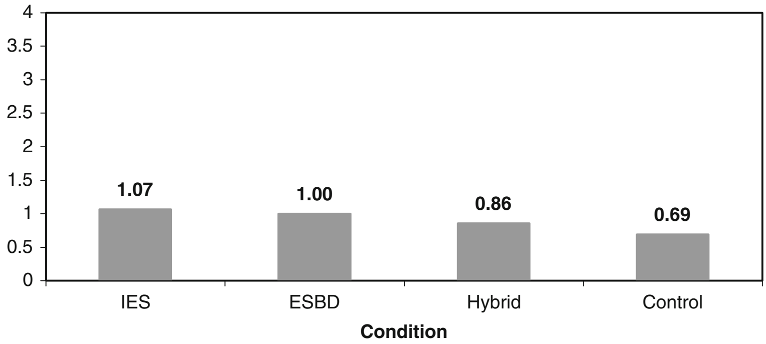

Fig. 3 Observer ratings of extent to which teachers elicited students' prior conceptions

Students in all three programs of professional development were judged by observers to be more likely than students in control classrooms to be able to provide an explanation for why they were engaged in a particular activity that was linked to a big idea in the Earth science unit (Fig. 2). At the same time, there were no statistically significant differences among the program designs with respect to students' understanding of why they were engaged in the day's instructional activities.

Teachers in all four conditions made little use of student preconceptions in class, and there were no differences between treatment and control conditions on this observation measure (Fig. 3). Further, there were no statistically significant differences among program designs with respect to teachers' elicitation of students' prior ideas about the concepts they were teaching that day.

Discussion and Conclusion

The pattern of results from the questionnaire data suggests that, overall, the designs differential effects on instructional planning were consistent with the differences among the designs' emphasis on planning units of instruction. After a year, teachers in the *ESBD* and *Hybrid* conditions reported significant changes to their unit planning process, a finding that is also consistent with the professional development designs for those conditions. In particular, teachers reported that the programs had affected both the process by which they planned and its content.

There were, to be sure, some differences between what was expected and what teachers experienced. For example, a majority of teachers in the *IES* condition did report after the workshop that they engaged in unit planning; only by examining data from the implementation questionnaire was it apparent what the nature of effects on unit planning were. These qualitative data present a picture consistent with the design, but they also suggest that even in a condition focused on curriculum adoption, many teachers engaged in some form of adaptation of the curriculum materials in planning for unit implementation. A separate analysis of control teachers' questionnaire data (Penuel and Gallagher 2008) revealed that they, too, had made some changes to their planning process since the beginning of the project, which confirms the ubiquity of adaptation in teachers' instructional planning.

The observational data indicated both areas where the professional development designs achieved their intended impacts, as well as areas where there were no significant impacts. With respect to the facet of interpretation, the two designs that had a significant impact on instruction focused more on teaching teachers explicitly about the facets as part of the professional development (as opposed to their being implicit in the design of the materials). All three designs produced students who could provide explanations for why their teacher had them engage in particular activities with reference to a big idea in the unit. At the same time, none of the designs had an impact on the probability that students would be observed engaging in explanation or application. Further, in contrast to the data on instructional planning that would suggest a greater attention to assessments, observers did not find teachers making use of preconceptions in their instruction.

Both the questionnaire and observations provide only a partial view on the impact that each design may have had. The questionnaire relied on self-report data; future studies should develop specific verbal protocols for studying teacher planning, which could be embedded into the planning process itself (e.g., Ericcson and Simon 1993). Resources limited the number of observations we could conduct in classrooms, making it possible that observers did not capture some impacts of the professional development. Nonetheless, the instructional planning data reports were consistent within condition, suggesting a uniform effect of each treatment. Further, the observation data focused on aspects of instruction that professional developers believed should be observable on a daily basis, regardless of where a teacher was in teaching their units. Thus, we believe that despite the limitations of the data, they provide valid, if preliminary, evidence of the impacts of the designs on teachers' instructional planning and enactment.

In sum, the study shows that professional development designs that aim to teach for understanding can differ, and with respect to instructional planning and enactment, these differences can make a difference. The differences in this particular study were greatest when comparing the designs' impacts on instructional planning. Those designs that emphasized those skills as part of their workshops had the greatest impact on planning. Differences in effectiveness of the designs on instructional practice were fewer, but there was evidence that with respect to requiring students interprets and uses analogies and models, the designs were not all equally effective.

A challenge remains to demonstrate that these "differences make a difference" in improving student achievement. We are exploring this question in our larger study, but it is important to recognize that we cannot answer easily the question of whether changes in teacher practice lead to changes in student learning. We cannot randomly assign teachers to have different *experiences* of professional development or enactment, and so we will never be sure that those experiences are the causes of changes to practice or teacher learning. Furthermore, there are lots of influences on student learning besides professional development and even teaching itself. Nonetheless, if results from survey measures of the kind we used in this study can be linked in correlational analyses to changes in teaching and learning, then researchers will likely see the utility of these kinds of measures in an even more positive light.

References

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education, 13*(1), 1–12. doi:10.1023/A:1015171124982.

Atkin, J. M., & Black, P. (2003). *Inside science education reform: A history of curricular and policy change*. New York: Teachers College Press.

Blumenfeld, P., Soloway, E., Marx, R. W., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3&4), 369–398. doi:10.1207/s15326985ep2603&4_8.

Bodilly, S. J. (1998). *Lessons from new American schools' scale-up phase*. Santa Monica: RAND.

Boone, W. J., & Kahle, K. B. (1998). Student perceptions of instruction, peer interest, and adult support for middle school science: Differences by race and gender. *Journal of Women and Minorities in Science and Engineering, 4*, 333–340.

Brown, J. L. (2004). *Making the most of understanding by design*. Washington, DC: Association for Supervision and Curriculum Development.

Bruner, J. S. (1960). *The process of education*. Cambridge: Harvard University Press.

Bybee, R. (1993). *Reforming science education: Social perspectives and personal reflections*. New York: Teachers College Press.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145–170. doi:10.3102/01623737023002145.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education, 77*(3), 211–244.

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

Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal, 44*(2), 298–338. doi:10.3102/0002831207302501.

Crawford, B. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching, 37*(9), 916–937. doi:10.1002/1098-2736(200011)37:9<916::AID-TEA4>3.0.CO;2-2.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.

Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal, 38*(4), 813–834. doi:10.3102/00028312038004813.

Davis, E. A., & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3–14. doi:10.3102/0013189X034003003.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81–112. doi:10.3102/01623737024002081.

Ericcson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge: MIT Press.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*(6), 643–658. doi:10.1016/S0742-051X(03)00059-3.

Gardner, H., & Dyson, V. (1994). Teaching for understanding in the disciplines and beyond. *Teachers College Record, 96*(2), 198–218.

Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945. doi:10.3102/00028312038004915.

Gess-Newsome, J. (1999). Secondary teachers' knowledge and beliefs about subject matter and their impact on instruction. In J. Gess-Newsome & L. M. Lederman (Eds.), *Pedagogical content knowledge and science education* (pp. 51–94). Boston: Kluwer.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3/4), 381–391. doi:10.1080/135406002100000512.

Haney, J. J., & Lumpe, A. T. (1995). A teacher professional development framework guided by reform policies, teachers' needs, and research. *Journal of Science Teacher Education, 6*(4), 1573–1847. doi:10.1007/BF02614642.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. doi:10.3102/00028312042002371.

Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teachers' classroom practices. *Journal of Research in Science Teaching, 42*(6), 668–690. doi:10.1002/tea.20069.

Johnson, C. C. (2007). Whole-school collaborative sustained professional development and science teacher change: Signs of progress. *Journal of Science Teacher Education, 18*(4), 1573–1847. doi:10.1007/s10972-007-9043-x.

Kubitskey, B., & Fishman, B. J. (2006). A role for professional development in sustainability: Linking the written curriculum to enactment. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (Vol. 1, pp. 363–369). Mahwah: Erlbaum.

Lieberman, A., & Miller, L. C. (2001). *Teachers caught in the action: Professional development that matters*. New York: Teachers College Press.

Linn, M. C., Songer, N. B., Lewis, E. L., & Stern, J. (1993). Using technology to teach thermodynamics: Achieving integrated understanding. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science* (pp. 5–60). New York: Springer.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345–375. doi:10.1146/annurev.psych.51.1.345.

Lotter, C., Harwood, W. S., & Bonner, J. J. (2006). Overcoming a learning bottleneck: Inquiry professional development for secondary science teachers. *Journal of Science Teacher Education, 17*(3), 185–216. doi:10.1007/s10972-005-9002-3.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics, 99*(5), 258–271.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks: Corwin Press.

Lumpe, A., Haney, J., & Czerniak, C. (2000). Assessing teachers' beliefs about their science teaching context. *Journal of Research in Science Teaching, 37*, 275–292. doi:10.1002/(SICI)1098-2736(200003)37:3<275::AID-TEA4>3.0.CO;2-2.

Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research in Science Teaching, 28*(1), 41–53. doi:10.1002/tea.3660280105.

Patton, M. Q. (1979). Evaluation of program implementation. In L. Sechrest (Ed.), *Evaluation Studies Review Annual* (Vol. 4). Newbury Park: Sage.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921–958.

Penuel, W. R., & Gallagher, L. P. (2008). *Comparing three approaches to preparing teachers to teach for deep understanding in Earth science: Short-term impacts on teachers and teaching practice*. Menlo Park: SRI International.

Penuel, W. R., & Means, B. (2004). Implementation variation and fidelity in an inquiry science program: An analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching, 41*(3), 294–315.

Penuel, W. R., Shear, L., Korbak, C., & Sparrow, E. (2005). The roles of regional partners in supporting an international Earth science education program. *Science Education, 89*(6), 956–979.

Radford, D. L. (1998). Transferring theory into practice: A model for professional development for science education reform. *Journal of Research in Science Teaching, 35*(1), 73–88. doi:10.1002/(SICI)1098-2736(199801)35:1<73::AID-TEA5>3.0.CO;2-K.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Rivet, A. (2006). Using transformative research to explore congruencies between science reform and urban schools. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 578–584). Mahwah: Erlbaum.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2004). *Evaluation: A systematic approach* (7th ed.). Newbury Park: Sage.

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal, 44*(2), 252–297. doi:10.3102/0002831207302498.

Scheirer, M. A. (1994). Designing and using process evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 40–68). San Francisco: Jossey-Bass.

Schneider, R. M., & Krajcik, J. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education, 13*(3), 221–245. doi:10.1023/A:1016569117024.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Singer, J. E., Krajcik, J., Marx, R. W., & Clay-Chambers, J. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist, 35*(3), 165–179. doi:10.1207/S15326985EP3503_3.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies, 31*, 143–175. doi:10.1080/002202799183205.

Spillane, J. P. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge: Harvard University Press.

Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitous pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record, 98*, 449–481.

Supovitz, J. A., & May, H. (2004). A study of the links between implementation and effectiveness of the America's Choice comprehensive school reform design. *Journal of Education for Students Placed at Risk JESPAR, 9*(4), 389–419. doi:10.1207/s15327671espr0904_4.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(2), 963–980. doi:10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0.

Treagust, D. F., Jacobowitz, R., Gallagher, J. L., & Parker, J. (2001). Using assessment as a guide in teaching for understanding: A case study of a middle school science class learning about sound. *Science Education, 85*(2), 137–157. doi:10.1002/1098-237X(200103)85:2<137::AID-SCE30>3.0.CO;2-B.

Tushnet, N. C., Millsap, M. A., Abdullah-Welsh, N., Brigham, N., Cooley, E., Elliott, J., et al. (2000). *Final report on the evaluation of the National Science Foundation's Instructional Materials Development Program*. San Francisco: WestEd.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge: Harvard University Press.

van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. *Journal of Research in Science Teaching, 38*(2), 137–158. doi:10.1002/1098-2736(200102)38:2<137::AID-TEA1001>3.0.CO;2-U.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.

Wiske, S. (1997). *Teaching for understanding: Linking research with practice*. San Francisco: Jossey-Bass.