THE JOURNAL OF THE LEARNING SCIENCES

ISSN: 1050-8406 (Print) 1532-7809 (Online) Journal homepage: https://www.tandfonline.com/loi/hlns20

Preparing Teachers to Design Instruction for Deep Understanding in Middle School Earth Science

William R. Penuel & Lawrence P. Gallagher

To cite this article: William R. Penuel & Lawrence P. Gallagher (2009) Preparing Teachers to Design Instruction for Deep Understanding in Middle School Earth Science, THE JOURNAL OF THE LEARNING SCIENCES, 18:4, 461-508, DOI: 10.1080/10508400903191904

To link to this article: https://doi.org/10.1080/10508400903191904

Published online: 28 Oct 2009.

THE JOURNAL OF THE LEARNING SCIENCES, 18: 461–508, 2009

ISSN: 1050-8406 print / 1532-7809 online
DOI: 10.1080/10508400903191904

Preparing Teachers to Design Instruction for Deep Understanding in Middle School Earth Science

William R. Penuel
SRI International
Menlo Park, California

Lawrence P. Gallagher
SRI International
Menlo Park, California

This study compared the efficacy of 3 approaches to professional development in middle school Earth science organized around the principles of Understanding by Design (Wiggins & McTighe, 1998) in a sample of 53 teachers from a large urban district. Teachers were randomly assigned to a control group or to 1 of 3 conditions that varied with respect to the conceptions of ideal curriculum use embedded within. Teachers either designed units of instruction, adopted units developed by expert Earth scientists and Earth science educators, or learned principled ways to adapt expert-designed curricula. Relying on data from surveys and independent ratings of naturally occurring teacher assignments, we used hierarchical linear modeling techniques to analyze the impacts of the professional development on how teachers planned and coordinated instruction. Our results suggest that to realize positive effects on both their planning and coordination of instruction, teachers need access to high-quality curriculum materials and professional development that helps them plan for principled adaptation of those materials.

For much of the past century, curriculum scholars did not view teachers as having a significant role to play in curriculum design. In the model of schooling that formed in

Correspondence should be addressed to William R. Penuel, Center for Technology in Learning, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025. E-mail: william.penuel@sri.com

the early 20th century and persists today in many schools and districts, teachers were and still are expected to implement, but not design, curriculum materials (Tyack & Cuban, 1995; Tyler, 1949). When the National Science Foundation began investing in the development of mathematics and science curricula during the post-Sputnik era, the dominant view expressed by leading figures in the early curriculum development projects was that "the best minds in any particular discipline must be put to work on the task" (Bruner, 1960, p. 19). Teachers could be involved in testing curricula, but few embraced the idea that teachers could contribute expertise to designing curricula (Atkin & Black, 2003). The belief was that it was "a great deal to ask" (Bruner, 1960, p. 68) of a teacher to have the knowledge of subject matter required to design curricula or even to implement them in a way that reflected designers' ideals.

The present era in science curriculum studies is characterized by more diversity in the roles teachers are expected to play in curriculum design. There still are calls for curricula that attempt to specify precisely what teachers should say and do in the classroom, on the assumption that teachers lack the necessary knowledge and skill to make design decisions on their own. At the same time, there is a growing recognition of the necessity and desirability of preparing teachers to make effective decisions in designing instructional experiences for students with curricula. Researchers who adopt this perspective view written curriculum materials not simply as designs that teachers implement with students; rather, teachers' activities to plan and enact curricula contexts represent a critical second step in curriculum design (Ben-Peretz, 1990; Connelly & Ben-Peretz, 1980). These decisions in turn are what shape curricular enactment and student opportunities for learning (Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2009; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Rivet, 2006).

This article reports results of a study comparing the efficacy of three professional development programs in developing teachers' skill in designing instruction. The three programs varied with respect to the roles teachers were expected to play in curriculum design and, consequently, in their approaches to preparing teachers to design instruction. In the first program, teachers learned principles of curriculum design and applied them to design units with available materials that addressed local standards (the *Design* condition). In the second program, teachers learned how to follow an inquiry-oriented curriculum unit designed for middle school Earth science developed by a team of learning scientists and subject matter experts (the *Adoption* condition). In the third program, teachers learned principles of curriculum design as they did in the first program, but they also had access to the curriculum materials teachers in the second program were expected to use. In the third program, teachers learned how to apply the principles of curriculum design to adapt materials to design units that fit their school's and district's priorities for learning (the *Principled Adaptation* condition).

Following Fishman, Marx, Best, and Tal (2003), we sought in our broader project to trace the path from teachers' experience of the professional development program to impacts on teaching and learning; in this article, we first present findings about teacher experiences and teacher learning. Next we present survey data about teachers' experience in their initial workshops to document the degree to which the professional development programs engaged them in ways congruent with those programs' design. Then, we present results of an analysis of naturally occurring teacher assignments, an analysis of teachers' actual lesson plans and associated materials (e.g., student worksheets, assessments, and scoring rubrics) intended to provide evidence of the impact of the programs on teachers' skill in designing instruction.

Researchers who have used teachers' naturally occurring assignments have reported that using such artifacts provides a way to measure the quality of students' opportunities to learn; those measurements can in turn be used to judge the success of programs (Borko et al., 2006). As our findings here illustrate, teachers' naturally occurring assignments also provide valuable insight into teachers' assumptions about what students should learn and how they should learn—beliefs that two of the professional development programs targeted for change.

BACKGROUND TO THE STUDY

Curriculum implementation in science is a problem that has long vexed policymakers, curriculum developers, and science educators. In the late 1950s, when the National Science Foundation first funded the design of hands-on science materials for schools, curriculum developers became frustrated by what they saw as teachers' failure to enact curricula in ways that reflected an understanding of the structure of scientific disciplines (Bruner, 1960). More recently, learning sciences researchers engaged in curriculum development projects have found that teachers using reform-oriented curriculum materials enact them only to a limited extent or in ways that do not reflect the intentions of designers (Brown & Campione, 1996; Reiser et al., 2000; Songer, Lee, & Kam, 2002; Spillane, 1999). These observations about science curriculum implementation are similar to those in evaluation studies about reading and mathematics programs, in which concerns about implementation and the consequences of poor implementation on student outcomes are recurring themes (Rowan, Camburn, & Correnti, 2004; Sarama, Clements, & Henry, 1998).

Studies of curriculum implementation frequently explain difficulties in implementation as teachers' failures to use curriculum materials as expert curriculum writers intended. For example, researchers who focus on fidelity of implementation investigate how well a curriculum or intervention is implemented in comparison with an ideal posited by researchers or curriculum developers (O'Donnell, 2008; Snyder, Bolin, & Zumwalt, 1992). Research with this focus seeks to address under what circumstances teachers use curriculum with fidelity and to identify strategies for improving fidelity (Remillard, 2005). Implementation fidelity research can and does explore variation in implementation, but the primary focus is on explaining variation with reference to program or curriculum designers' ideals (e.g., Penuel, Fishman et al. 2007; Penuel & Means, 2004).

An important criticism of studying curriculum use from a fidelity perspective is that such research often fails to take into account the work teachers must do to adapt curricula to their particular classrooms. McLaughlin (1976), an early critic of a fidelity perspective, noted that adapting curricula to the local context is not only necessary but also desirable to make curricular materials effective for teachers' students. Connelly and Clandin (1988) attributed the failures of early curriculum reform efforts funded by the National Science Foundation to the failure of curriculum and staff developers to consider the experience of teachers and students in their curriculum situations (see also Randi & Corno, 1997). More recently, learning scientists have pointed out that the local classroom context has primacy when it comes to curriculum context (Squire, MaKinster, Barnett, Luehmann, & Barab, 2003). In addition, accountability pressures weigh heavily in teachers' decisions about how much time to give to particular science topics and about how to adapt curricular materials to their classrooms (Li, Klahr, & Siler, 2006; Marx & Harris, 2006). The adaptations teachers make to curricula can occur before, during, and after enactment (Davis & Varma, 2008), and those adaptations may be inspired by particular classroom situations, when students and teachers interact with materials or discuss an idea (Marton et al., 2004).

From a policy perspective, careful consideration of the work teachers do to adapt curriculum materials as part of planning and enacting instruction is fundamental, given that most states still leave decisions about what curriculum materials and instructional strategies to use to local districts and often to teachers as well, even in this era of high-stakes accountability (Ingersoll, 2003). Furthermore, most teachers have access to a wide array of instructional materials offered by different publishers in the private sector that reflect varying and sometimes competing curricular purposes (Meyer & Rowan, 2006). Acknowledging this reality, some scholars have argued that teachers need to have a significant role in curriculum design, because teachers, not outside experts, are the actors in the school system with primary responsibility for student learning (Keys & Bryan, 2001; Parke & Coble, 1997).

As agents of change, however, teachers may not have access to the resources and expertise they need to make effective adaptations to curricula. Numerous studies have shown that teachers confronted with the challenge of enacting ambitious new curricula whose purposes and activities diverge from teachers' current practice vary in their approaches to designing instructional experiences for students using those curricula (Reiser et al., 2000; Remillard, 1999; Sherin, 2002). When teachers do not have adequate professional development and support to make use of curricular materials, they may not implement the materials at all or may do so in ways that limit student opportunities to learn from those materials (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004; Pea & Collins, 2008). As part of professional development, teachers may need specific guidance to build their skill in making productive adaptations of curricular materials that fit their contexts and that are responsive to the interests and needs of their students (Brown, 2008; Brown & Edelson, 2003).

Planning and Enacting Instruction as a Phase of Curriculum Design

In her review of different approaches to studying teachers' use of mathematics curricula, Remillard (2005) argued that curriculum use is best understood as a process of teachers' participating with, rather than following or subverting, curricular aims and materials. Her concept views curriculum resources as cultural artifacts (Wartofsky, 1973) or mediational means (Wertsch, 1991, 1998) that are products of particular social and historical institutions. In adapting these materials to meet local standards and to align them with student interests and needs, teachers populate the materials with their own intentions, thereby making them partially their own and in the process transforming how the materials affect student learning (Brown, 2002). At the same time, use of materials may also change teachers themselves, because using curriculum materials may enable new arrangements for learning and forms of classroom interaction to emerge that the teacher had not imagined (Collopy, 2003; Drake & Sherin, 2006).

From this participatory perspective on curriculum use, important questions remain about which factors influence teachers' relationship with curriculum resources. Empirical studies of curriculum use suggest that teachers' knowledge, attitudes, and experience all affect how they engage with curriculum materials. For example, teacher knowledge and classroom logistics concerns influence teachers' planning decisions (McCutcheon, 1980) and management of classroom discussion (Carlsen, 1991, 1993). In their case studies of two urban elementary school teachers, Drake and Sherin (2006) found that teachers' early experiences with mathematics and their beliefs about themselves as mathematics learners shaped patterns of curriculum adaptation. The variability in curriculum use that Drake and Sherin reported echoes the findings of other studies of curriculum use (Collopy, 2003; Penuel, Fishman, et al., in press; Remillard, 1999).

Other studies have suggested that teachers vary with respect to their *pedagogical design capacity* (Brown, 2008; Brown & Edelson, 2003), that is, teachers' skill in being able to design instructional experiences for students by adapting curriculum materials to their classrooms. For example, some teachers do not view intentional curriculum adaptation as a "legitimate" or "high-priority" part of their job (Bullough, 1992). Other teachers report that the process of analyzing curriculum as part of the planning process is destabilizing and troubling (Crawford, Korbak, & Gong, 2003; Schwarz et al., 2008). Researchers have found that in projects that involve teachers in curriculum design or codesign, teachers often have trouble thinking as designers do (Reiser et al., 2000), and adaptations of materials that are inconsistent with the design team's goals are still possible (Penuel & Yarnall, 2005).

Taken together, these findings are troubling for a number of reasons. To the extent that successful teaching engages teachers' knowledge of curricular purposes (Shulman, 1986), teachers need to have a good understanding of the connections among disciplinary ideas and an ability to analyze the likely adequacy of particular curriculum materials for developing student understanding of those ideas. Most reform curricula require new, unfamiliar roles for students and teachers in the classroom (Crawford, 2000; Songer et al., 2002) and an adaptive style of teaching that is responsive to how students learn (Sherin, 2002). In addition, today's science curricula seek greater coherence in their focus on core scientific ideas and the connections among them (Kali, Linn, & Roseman, 2008). Without outside assistance, teachers may not be able to discern the design rationale for particular activities (Davis & Varma, 2008), understand the roles expected of them and their students (Spillane & Jennings, 1997; Tushnet et al., 2000), or distinguish coherent from incoherent sequences of activities (Lin, 2008; Lin & Fishman, 2004).

Developing Teachers' Capacity to Design Instruction With Curriculum

Two key features characterize some models of professional development aimed at building teachers' design capacity. First, some models incorporate the idea that the materials themselves should be learning resources for teachers (Ball & Cohen, 1996; Davis & Krajcik, 2005; Schneider & Krajcik, 2002). These models are based on the premises that teachers need extra supports to develop the knowledge required to teach particular subjects (Davis & Krajcik, 2005) and that high-quality materials make it easier to plan productive curriculum experiences (Connelly & Clandin, 1988). Second, some models explicitly provide opportunities for teachers to collaborate with peers and interact with expert mentors. Other teachers are an important resource to consider in professional development designs in terms of adapting curricula, given that they may have a deep knowledge of the constraints and opportunities of particular school contexts (Judson & Lawson, 2007; Lang, Drake, & Olson, 2006). Interactions with colleagues can also help teachers make their interpretations and instructional decisions more explicit and visible to others (Remillard, 2005). Mentors can help teachers address gaps between curriculum writers' design intentions and teacher enactment (Ball & Cohen, 1996; Davis & Varma, 2008).

An illustration of how these different model components can support teachers' developing design skills is provided by professional development supports designed to support the Investigating and Questioning our World Through Science and Technology (IQWST) science curriculum materials developed by researchers at the University of Michigan and Northwestern University (Krajcik & Reiser, 2004). The goal of professional development for IQWST is to prepare teachers to enact the curriculum in ways that are congruent with researchers' design intentions while adapting it to their classroom contexts (Fishman et al., 2003). To accomplish this goal, professional developers use a cycle of planning, enacting, and reflecting called CERA, which stands for Collaborative construction of understanding, Enactment of new practices, Reflection on practice, and Adaptation of materials and practices (Marx, Blumenfeld, Krajcik, & Soloway, 1998).

Over the years, University of Michigan researchers have developed materials and processes to support the CERA process with the IQWST materials. For example, researchers have designed features for the teachers' guides to educate teachers about the scientific concepts they are teaching and to help them respond to typical student difficulties in understanding those concepts (Davis & Krajcik, 2005). In addition, Lin (2008) developed and tested the Planning, Enactment, and Reflection Tool (PERT), which uses scaffolding that helps teachers determine and reflect on how their curriculum modification decisions affect the designer-specified interconnections between content and inquiry standards within and across different lessons in a curriculum unit. PERT focuses explicitly on allowing teachers to reflect on how adaptations affect coverage of standards and connections among ideas. Lin found that teachers' use of these scaffolds significantly improved their understanding of underlying unit structures.

In recent years, empirical research has also emerged to investigate the efficacy of different approaches to developing teachers' capacity to design instruction with high-quality materials. Some studies have found that presenting teachers with curriculum design principles can help them attend to core features of particular frameworks for teaching and learning (Bybee, 1997; Edelson, 2001; Schwarz & Gwekwerere, 2007). In addition, Fishman and colleagues (2003) were able to trace how changed curriculum and professional development features affect teacher and student learning in small numbers of classrooms. However, no larger scale studies have been undertaken of the impact on teacher practice and student learning of programs aimed specifically at developing teachers' capacity to design instruction in science. Moreover, studies *comparing* the efficacy of different approaches to building teachers' pedagogical design capacity within a single context have not been conducted. Such studies are important, given the many possible approaches to building this capacity, each of which confers potentially different benefits and costs.

Linking Professional Development to Teacher Learning in Learning Sciences Research

The ultimate task of studies of the effectiveness of professional development is to determine whether a link exists between professional development and student achievement (Loucks-Horsley & Matsumoto, 1999). But the impact of any professional development on student learning is likely to be indirect: It will be a function of the design of the program, the enactment of the design, and the design's impact on teachers' cognition and their instructional practice (Fishman et al., 2003; Wayne, Yoon, Zhu, Cronen, & Garet, 2008). To link professional development to learning outcomes, it is first necessary to establish teachers' experiences in professional development and determine what they can articulate about what they have learned from those experiences. Then, researchers can analyze whether and how the programs also directly affect teachers' knowledge and skill, students' opportunities to learn, and students' achievements.

Consider first how designs might vary with respect to professional development aimed at developing teachers' skill in designing instruction. Some programs might pay scant attention to teachers' role in designing instruction. We can extend Remillard's (2005) framework, designed to apply to studies of curriculum use, to describe professional development models by saying that these programs adhere to a "following or subverting" concept of curriculum use, expecting teachers to follow curricula and avoid adapting activities altogether. Professional development would be likely to include activities intended to improve the fidelity of implementation, such as providing teachers with explicit instructional guidance about what to say or do in class. Other programs might focus explicitly on building teachers' skills in designing instruction using a variety of curriculum materials. Those programs would be inspired by the "drawing on" conception of curriculum use—developing specific strategies to teach teachers how to design powerful and engaging instructional experiences for students. Still other programs might seek ways to help teachers make principled adaptations of curriculum materials, consistent with a "participating with" conception of curriculum use. In such programs, professional development activities might be designed to help teachers select materials that best address local standards and that are appropriate for their students. Such programs might also provide teachers with strategies to elicit students' prior knowledge and interest in particular topics and to adapt activities to connect scientific ideas accordingly.

Just as designs vary, teachers' experiences of professional development can also vary as a result of leaders' enactment of designs and of the different sets of knowledge and beliefs the teachers bring to the learning situation (Spillane, 1999; Wayne et al., 2008). For example, different groups of professional developers may adopt different designs for introducing teachers to the same curriculum materials (Penuel, Fishman et al., 2007). For their part, teachers participating in collaborative professional development may learn different things from their experiences, depending on their prior knowledge and beliefs about teaching (Johnson, 2007). In this connection, incorporating interactions designed to provide feedback to professional developers on teachers' evolving understanding of curricular purposes and activities may be critical for designing experiences that promote teacher learning (van Driel, Beijaard, & Verloop, 2001). In terms of gauging differences among teachers with respect to experience, teachers themselves are key informants: What they say provides critical clues about the meaning they derive from their professional development experiences.

Measuring the impact of professional development on teaching practice in science has proven challenging to researchers, particularly in larger scale studies. Case studies have allowed for fine-grained analyses that trace the footprint of teacher learning designs on teaching practice (e.g., Davis, 2008; Fishman et al., 2003). But studies of multiple teachers have tended to rely extensively on survey self-report measures to estimate the impacts of professional development on instruction (e.g., Garet, Porter, Desimone, Birman, & Yoon, 2001; Supovitz & Turner, 2000). Such approaches potentially introduce bias into estimates of impact, particularly when survey questions use language to describe practices that are similar to those in teachers' professional development (Schwartz, 1999; Schwartz & Oyserman, 2001). In addition, self-report measures have proven inadequate in the past for making judgments about the quality of instruction (Porter, 2002).

A promising approach to measuring the quality of instruction—developed initially for other school subjects—is to collect and analyze naturally occurring teacher assignments. In reading and mathematics, researchers collecting such assignments have used rubrics to characterize them with respect to their intellectual rigor and authenticity (Newmann, Lopez, & Bryk, 1998; Wehlage, Newmann, & Secada, 1996). Validation studies have found that teacher assignments can be rated reliably and with consistency and that the quality of assignments correlates significantly with direct observations of instructional quality (Clare & Aschbacher, 2001). Validation studies have also demonstrated positive associations between the quality of assignments and the work students produce in class (Matsumura, Patthey-Chavez, Valdez, & Garnier, 2002; Shkolnik et al., 2007) and have linked higher scores on assignments to higher student achievement (Newmann, Bryk, & Nagaoka, 2001).

A recent study of classroom artifacts in science yielded findings that suggest that analyzing teacher assignments may be a useful method for studying instructional quality in science classrooms. In that study, Borko and colleagues (2006) had teachers "scoop" artifacts such as lesson plans, handouts, samples of student work, and grading rubrics into notebooks; researchers then scored the notebooks on 11 dimensions and compared the resulting scores with ratings made by independent observers on the same dimensions. The independent raters' scores for the notebooks were similar to those based on observation; in addition, Borko and colleagues (2006) found that the way in which science instruction was portrayed was similar across notebooks and observations.

At the same time, Borko and colleagues (2006) cautioned that artifact analyses may not yield reliable enough estimates of individual teachers' practice for use in judging that performance. In their study, agreement on exact ratings of notebook scores across dimensions was low (between 22% and 47%), and agreement within 1 point on a 5-point scale for the notebook rubrics was modest (between 75% and 91%). Furthermore, the researchers noted that interrater agreement was lower for notebooks in which teachers' reflections on their goals and context for the lessons submitted were limited, leaving raters with little evidence to judge lessons on some dimensions. As a result, Borko and colleagues (2006) concluded that notebooks were not yet reliable enough as sources of data on instructional quality to support a high-stakes personnel evaluation; however, they argued that their validity analyses suggested that notebooks could contribute evidence to studies documenting the impacts of instructional reforms.

Our own perspective, for which we provide evidence in this article, is that naturally occurring teacher assignments can be used to compare the efficacy of different approaches to developing teachers' pedagogical design capacity. We do not believe that assignments alone provide sufficient evidence for judging instructional quality, particularly those dimensions of instructional quality that pertain to how teachers make adaptations to curricula on the fly in response to their students' beliefs, interests, and level of understanding. Rather, teachers' assignments provide valuable insight into their invention and appropriation of tasks to meet particular learning goals. The tasks teachers decide to pose to students are reflections of multiple concerns they have about teaching, such as what they value most and consider worthy of understanding, their perceptions of the constraints of their classrooms and the capabilities of their students, and the possibilities they see in particular curricula (Remillard, 1999; Rivet, 2006; Rivet & Krajcik, 2004). Although researchers cannot easily distinguish these different influences by analyzing assignments in isolation, they can begin to trace the path from professional development to impact when assignments are coupled with data from teachers' own perspectives on their planning process and the professional development.

Determining the potential for naturally occurring teacher assignments to yield information on what Remillard (1999) called the *design arena* of curriculum use has been fundamental to our research. The design arena as Remillard (1999) defined it pertains to teachers' activities to invent or appropriate particular tasks for students. Design is only one arena of curriculum use, because use also encompasses adaptations during the context of enactment, but design is important for measuring teachers' pedagogical design capacity. It is particularly well suited to measuring the impacts of the programs of professional development that are the focus of our study, given that each of these programs aimed to affect the kinds of tasks teachers present to students. In particular, all three programs we describe here sought to increase the number of tasks presented to students that required them to interpret information, develop explanations, and apply knowledge of basic science facts to solving challenging problems (Wiggins & McTighe, 1998). At the same time, the programs varied with respect to the extent to which each expected teachers to design, adopt, or adapt particular tasks for instruction; thus, using naturally occurring assignments as a measure of pedagogical design capacity helps illuminate which of these strategies is likely to be most effective in developing teachers' skill in designing instruction.

COMPARING THE EFFICACY OF THREE APPROACHES TO DEVELOPING TEACHERS' DESIGN CAPACITY

This article reports results from an efficacy study that took place in a single large urban district in the southeastern United States. In an efficacy study, an intervention or set of interventions is tested under relatively good or even ideal conditions to investigate its potential impact and key implementation characteristics (Flay et al., 2005; Glascow, Lichtenstein, & Marcus, 2003). The conditions in the participating district were ideal for a study of professional development focused on teaching for understanding for two reasons: (a) The district had recently adopted an Understanding by Design approach (Wiggins & McTighe, 1998) for its teachers to use in guiding the development during the school year of four 9-week units, each focused on developing deep understanding of a few concepts; and (b) the district had not yet invested significant professional development in preparing teachers to use the approach. That context allowed us (a) to compare the efficacy of different approaches to preparing teachers to provide instruction designed to foster deep understanding and also (b) to test the efficacy of these approaches with a comparison group of teachers who, although they would not be exposed to the same professional development, would nonetheless be expected to teach to the same set of standards.

In the following sections, we describe the three professional development designs, the sample and measures used in our study, teachers' experiences of professional development, and impacts on planning and enactment of the professional development. Our goal was to answer the following questions:

1. How well are the professional development designs reflected in teachers' experiences of their professional development?
2. What are the impacts of each approach on teachers' instructional planning process?
3. What are the impacts of each approach on the quality of teachers' assignments?
4. What are the impacts of each approach on the quality of teachers' culminating performance assessments of student learning?

Description of the Three Professional Development Program Designs

All three professional development designs shared three elements that made this efficacy study possible. First, the designs were enacted within a single school district. Enacting the designs in one district meant that all teachers in the study were expected to teach to the same standards (i.e., those that applied to their grade level). Furthermore, enactment in a single district enabled leaders of the professional development to tailor their activities to the district's goals and priorities. Tailoring activities to local goals was important to our study design because past studies have shown a relationship between changes in teacher knowledge and practice and teacher perceptions about the coherence of professional development with local goals (Garet et al., 2001; Penuel, Fishman et al., 2007). Second, the teachers in all three programs were expected to plan and enact a 9-week unit in Earth science. The unit approach allowed us to collect data from all teachers at regular 3-week intervals. Third, the duration of professional development was the same for all programs. Each teacher participated in a 2-week summer workshop in the year before data collection began; during the following academic year, teachers participated in four 1-day workshops to further elaborate on their unit plans and to discuss enactment of their units. Holding duration constant across programs was important to the study design given that past research has demonstrated links between the length of time teachers spend in professional development and how much they learn from it (Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet et al., 2001; Supovitz & Turner, 2000).

Apart from these similarities, the programs differed with respect to their goals, the curriculum materials that teachers were expected and allowed to use, underlying conceptions of curriculum use, and the expected roles of teachers in curriculum use. Table 1 summarizes these points of variation, and we describe them in greater depth here. To read about how these programs reflected more general principles of effective professional development design, see Penuel, Benbow, and colleagues (2009).

The curriculum *Adoption* program concerned professional development to prepare teachers to enact an inquiry-based curriculum, *Investigating Earth Systems* (IES). The program for professional development and the curriculum modules were developed by the American Geological Institute (AGI) with funding from the National Science Foundation (NSF). AGI staff developers adapted their program of professional development to prepare teachers to implement specific modules that fit the school district's middle school Earth science standards. In the district in which we conducted the study, AGI worked with district leaders to select the four modules that were most closely aligned to the state standards: Dynamic Planet (sixth grade), Rocks and Landforms (sixth and seventh grades), Water as a Resource (seventh grade), and Astronomy (eighth grade).

For our study, we asked participants in the Adoption program to limit their use of materials to the IES curriculum as part of their units. The inquiry-based Earth systems science curriculum consists of a student edition with investigations and content and a teacher's edition with science background, indication of typical student misconceptions, teaching tips, materials management advice, assessments,

TABLE 1
Comparison of the Three Professional Development Program Designs

Variable	*IES (Adoption)*	*ESBD (Design)*	*Hybrid (Principled Adaptation)*
Goal of professional development	Prepare teachers to adopt IES units	Prepare teachers to design their own Earth science units	Prepare teachers to adapt IES units
Science content	Relevant district standards for structure of the Earth (Grade 6), water cycle (Grade 7), astronomy (Grade 8)	Relevant district standards for structure of the Earth (Grade 6), water cycle (Grade 7), astronomy (Grade 8)	Relevant district standards for structure of the Earth (Grade 6), water cycle (Grade 7), astronomy (Grade 8)
Duration of professional development	2-week workshop, plus 4 follow-up days with teachers during the school year (112 hr)	2-week workshop, plus 4 follow-up days with teachers during the school year (112 hr)	2-week workshop, plus 4 follow-up days with teachers during the school year (112 hr)
Curriculum materials	IES	Teachers select materials, which can include textbook and teacher-designed activities	IES, plus teacher-selected or teacher-designed materials
Conception of curriculum use[a]	Following or subverting	Drawing on	Participating with
Teacher's expected role in curriculum use[a]	Enactor of planned curriculum	Active designer of the enacted curriculum	Collaborator with curriculum materials to design enacted curriculum

Note. IES = Investigating Earth Systems; ESBD = Earth Science by Design.
[a]Adapted from Remillard (2005).

alignment with National Science Education Standards, and online teaching resources. IES emphasizes the importance of learning science through inquiry. Extended reading passages are few in the student editions, and those that are included are intended to help contextualize what students are learning from their investigations in class. All of the investigations require some tailoring on the part of teachers and students, tailoring that is expected to reflect issues of local concern in some cases (e.g., studying water use in students' homes) and student interests and questions in other instances. The investigations are also intended to help students gain an appreciation of the discovery process in Earth science, including challenges that have vexed scientists for decades (e.g., how to study the Earth's interior).

Most of the activities of the professional development aimed to prepare teachers to enact the investigations that are the heart of every module. The first part of the workshop covered topics that underpin the curriculum: typical module structure, the nature of inquiry-based science and the Earth systems approach, management of materials and students working in collaborative groups, teacher support, and the Web site and assessment components used in IES. In the second part of the summer workshops, teachers worked in groups to practice with activities as if they were students and to study the content of the IES modules they would use with their students. In the four follow-up training sessions, AGI staff met with the teachers to discuss issues and successes they experienced during the implementation. At these meetings, teachers also shared student work and assessments and discussed adaptations they had made to accommodate their students' ability levels.

The version of professional development AGI staff provided to teachers in this study corresponds closely to the "following or subverting" concept of curriculum use in Remillard's (2005) framework. AGI provided specific guidance to teachers for enacting all of the investigations, in the order they were presented in the modules, for the 9-week Earth science unit. AGI staff also asked teachers to refrain from using any other material, except material on the IES Web site.

The curriculum *Design* condition was the *Earth Science by Design* (ESBD) program. ESBD's goal is to prepare teachers to design their own Earth science units. Its name suggests its inspiration, namely the method of curriculum design called Understanding by Design (Wiggins & McTighe, 1998). Curriculum writers at TERC and AGI collaborated to develop the program with funding from NSF's Teacher Enhancement program. Its specific objectives were to teach for deeper understanding by focusing on "big ideas" and using an "Earth-as-a-system" approach; to design and apply appropriate assessment techniques, such as preconception surveys and authentic performance measures; and to use visualizations and satellite imagery to promote student understanding.

Teachers in the ESBD professional development program did not receive curriculum materials as part of their participation. Instead, those completing the program reorganized existing curricular materials, such as those from their textbooks or those they had developed themselves or collected from colleagues at professional conferences, into coherent units of instruction that targeted essential questions and enduring understandings and that culminated with a performance assessment. To prepare for the task of designing units, teachers in the ESBD workshop engaged in activities and discussions to consider the nature of understanding, to struggle with what was worthy of understanding, and to begin to comprehend the Earth-as-a-system approach to Earth system science. They learned the "backward design" process and practiced constructing a unit using the ESBD online unit planner. ESBD teachers also practiced developing assessments of student learning intended to measure students' level of understanding of big ideas. As part of the workshop, leaders gave teachers time 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 to reveal students' understandings (and misunderstandings) of the unit. Last, by considering the lessons they had used when they had 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. The workshop's sole content requirement was the expectation that teachers incorporate visualizations and Internet resources into their units.

Using Remillard's (2005) framework for analyzing conceptions of curriculum use and teachers' role in curriculum, ESBD corresponds most closely to the "drawing on" conception and casts teachers in the role of designers of instructional experiences. Teachers are to draw on existing resources they have ready-to-hand and may even have used before while reorganizing their units. In undertaking this task, however, the teacher acts as chief designer of the instructional experience for students. The ESBD program tries to prepare teachers for this role by giving them a framework for and practice in designing their own units in the context of an extended professional development experience.

We refer to the *Principled Adaptation* professional development program as the *Hybrid* program because it represents a blend of the ESBD and IES programs. Its goal was to prepare teachers to adapt IES units, using the Understanding by Design approach emphasized in ESBD. Its objectives encompassed the objectives of ESBD, and it also sought to prepare teachers by having them use IES investigations in designing their 9-week Earth science units.

For our study, we asked teachers assigned to the Hybrid program to use a combination of curriculum materials that included the IES materials. Teachers could use any combination of textbook materials, activities created by themselves or colleagues, and the IES investigations. To ensure at least some use of the IES materials, we asked that teachers use at least 50% of the IES investigations relevant to their grade level in constructing their units.

The activities of the professional development workshop included activities from both the ESBD and IES workshops. Like the teachers in the ESBD workshop, teachers in the Hybrid program engaged in activities and discussions about the nature of understanding and what is worthy of understanding. They also learned about the Earth-as-a-system approach to Earth system science and about the process of backward design. They 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 creating assessments of student learning intended to develop students' 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.

A key activity in the Hybrid workshop was analyzing or "unpacking" the IES unit structure and activities in terms of how they reflected Understanding by Design principles. Teachers reviewed student editions and identified the targeted enduring student understandings and essential questions that were to guide the investigations. They analyzed the features, including the culminating and interim assessments intended to measure student understanding, particularly with respect to the importance of assessments having an authentic, external audience and an ability to capture students' skill in integrating and synthesizing what they learned. Finally, teachers also identified strategies that the curriculum modules incorporated for engaging students in cognitively complex activities of explanation, interpretation, and application of knowledge.

The Hybrid program reflects a conception of curriculum use as teachers "participating with" materials. The assumption behind its approach to preparing teachers to adapt IES units is that teachers draw on curriculum materials to design the enacted curriculum. Teachers' own design activities are essential, in that teachers must always adapt materials to fit their local context. Behind program activities to prepare teachers for those design activities lies the assumption that teachers may vary in their preparation for engaging in the activities. The Hybrid program explicitly seeks to build teachers' pedagogical design capacity but does so in the context of working with high-quality curriculum materials developed by experts knowledgeable about inquiry science.

STUDY DESIGN

This study used an experimental design, in which teachers were randomly assigned to one of the three treatment conditions or to the Control condition. Random assignment studies pose the fewest threats to internal validity and, compared with other designs, are more likely to yield unbiased estimates of potential impacts (Shadish, Cook, & Campbell, 2002). Because the random assignment process took place after teachers in the study volunteered for professional development, study findings cannot be generalized beyond those groups of teachers. Other efficacy studies that study the impact of the interventions when teachers are compelled to participate are needed to determine potential under those conditions.

Statistically, ideal random assignment results in every teacher having an equal probability of being assigned to any condition (Shadish et al., 2002). At the same time, to minimize possible school effects, we sought to balance conditions within schools and grade levels as much as possible. Our randomization scheme had three key features. First, each teacher had exactly a 1 in 4 chance of being assigned to one of the three professional development programs or to the Control condition, thus meeting a fundamental condition of random assignment. Second, the number of teachers in each condition was balanced across the entire sample. Third, the widest possible dispersion of assignments to condition was made both within a school and within a single grade level in that school. The risk of cross-contamination of conditions within schools was extremely low; in 15 of the 19 participating schools there were never two teachers in the same condition in any grade, whereas in the remaining 4 schools, teachers in the same condition were teaching Earth science topics in different grades (and hence using different materials). For a more detailed description of the procedures used to assign teachers to condition, see Penuel and Gallagher (2006).

Participants

A total of 53 sixth-, seventh-, and eighth-grade teachers from 19 middle schools in a large urban district participated. Of these teachers, 13 were assigned to a program designed to prepare teachers to adopt a high-quality Earth science unit (the IES condition), 13 to a program designed to prepare teachers to design their own units (the ESBD condition), 13 to a program designed to prepare teachers to adapt high-quality Earth science units in a principled way (the Hybrid condition), and 14 to the Control group (the condition in which teachers did not receive professional development). There were no significant differences among groups with regard to any of the teacher characteristics presented in Table 2. Although on average teachers in the ESBD condition had less teaching experience than teachers in the other conditions, the differences among groups were not statistically significant for overall years of teaching and were marginally nonsignificant ($\chi^2 = 7.59$, $p = .06$)[1] for years of teaching science. As can be seen in Table 2, teachers in the Hybrid condition had on average less experience teaching science than teachers in the other three conditions.

As Table 3 shows, although members of every group had received some professional development in the Understanding by Design approach, the groups did not differ significantly in terms of overall exposure. Similarly, there were no significant differences among groups in professional development related to the development of classroom assessments or unit planning.

Sources of Data

Postworkshop Teacher Questionnaire

We collected data on teachers' experiences of their professional development through a researcher-administered questionnaire at the end of each workshop. We analyzed individual items taken from two indices from an earlier study (Garet et al., 2001) of effective professional development related to forms of engagement and interactions with workshop leaders and peers. For both sets of items, we examined which items a majority of teachers said had characterized their particular workshop. We also used Garet and colleagues' scale for comparing perceptions of coherence among teachers assigned to different conditions. In the original study,

[1]Chi-square tests with three degrees of freedom based on the null hypothesis that expected outcome is equal across all four conditions when a two-level random intercepts hierarchical linear model is fit (teachers nested within schools).

TABLE 2
Characteristics of Faculty Respondents to Questionnaire by Condition

Characteristic	*IES*	*ESBD*	*Hybrid*	*Control*
Gender, %				
Male	23	38	23	29
Female	77	62	77	71
Race/ethnicity[a], %				
White	77	46	46	79
Black	15	46	46	14
Hispanic/Latino	8	0	15	7
Asian	8	0	15	0
Other/unknown	0	8	8	0
Teaching experience, years				
Teaching	$M = 13.2$	$M = 14.9$	$M = 8.4$	$M = 12.8$
	$SD = 11.8$	$SD = 11.8$	$SD = 8.4$	$SD = 9.0$
Teaching science	$M = 11.5$	$M = 8.8$	$M = 4.3$	$M = 10.4$
	$SD = 9.6$	$SD = 6.1$	$SD = 3.3$	$SD = 8.1$
Highest degree[b], %				
Bachelor's	69	77	85	64
Master's	23	23	8	36
Educational specialist	8	0	8	0
Missing	0	0	0	0
Teaching assignment				
Grade 6	5	7	6	3
Grade 7	3	2	4	4
Grade 8	5	4	3	7

Note. The table is based on 53 teachers who submitted examples of assignments for analysis. IES = Investigating Earth Systems; ESBD = Earth Science by Design.
[a]Teachers could select multiple categories.
[b]Totals may not equal 100% because of rounding.

the reliability of this scale was high ($\alpha = .86$). To examine the focus on inquiry, we included an item asking teachers to report how much emphasis the workshop gave to scientific inquiry (0 = *none at all*, 3 = *a lot of emphasis*).

Postunit Implementation Teacher Questionnaire

The questionnaire we used to collect data on teachers' reported changes to instructional planning focused on a range of topics related to teachers' instructional planning process and unit implementation. Our analysis focused on the three aspects of the planning process.

Reported influence on the instructional planning process. This measure consisted of a single item related to the influence of the project. Teachers answered

TABLE 3
Participation in Professional Development, by Topic and Condition
(Percentage of Teachers)

Topic/Condition	*0 Hr*	*1–8 Hr*	*9–16 Hr*	*17–25 Hr*	*25+ Hr*
Understanding by Design					
IES	31	38	8	8	15
ESBD	23	69	0	8	0
Hybrid	31	54	8	0	8
Control	36	29	14	14	7
Total	30	47	8	8	8
Classroom assessment					
IES	15	31	23	0	31
ESBD	8	38	15	15	23
Hybrid	15	62	8	8	8
Control	36	29	7	21	7
Total	19	40	13	11	17
Unit planning					
IES	46	23	15	0	15
ESBD	15	38	15	23	8
Hybrid	31	46	8	0	15
Control	50	36	0	7	7
Total	36	36	9	8	11

Note. Percentages may not equal 100% across columns because of rounding. IES = Investigating Earth Systems; ESBD = Earth Science by Design.

the question "Has your experience in this project influenced how you plan instruction in science?" The three ordinal response options were 1 (*not at all*), 2 (*somewhat*), and 3 (*very much*).

Reported changes in how teachers organize assignments. This measure comprised four items related to teachers' instructional planning process as targeted by professional development in the Understanding by Design model: (a) use of activities aligned to big ideas of the discipline, (b) number of topics covered, (c) selection of materials to develop deep understanding, and (d) selection of materials intended to catch students' interest. Using a scale from 1 (*not at all*) to 3 (*very much*), teachers rated how much each item had changed since the beginning of the research project. A scale representing the raw sum of responses was constructed from the items, with a reliability of $\alpha = .81$.

Reported use of instructional materials. This measure comprised five items, each analyzed separately, focused on changes to teachers' use of materials when developing and teaching their Earth science units. Teachers responded to questions about their use of the district-selected textbook, the Internet, visualizations (a focus of the ESBD and Hybrid professional development), materials they created, and materials created by colleagues. For each, teachers responded on a 6-point scale ranging from *used much less* to *used a lot more*.

Teacher background in Earth science. As a measure of teachers' prior knowledge in Earth science, we asked teachers questions about their certification status in science, past professional development in Earth science topics, and undergraduate and graduate coursework in the field. Each of these variables served as control variables in the preliminary models we fit to the data on teacher assignment quality.

Teacher Assignments

The study team developed five 4-point rubrics to analyze naturally occurring teacher assignments. The rubrics for assignments addressed each of the following dimensions of assignments.

Scientific communication. This rubric pertained to the extent to which the assignment required students to engage in scientific communication to demonstrate their understanding of science content.

Indicators: Students were asked to engage with and communicate about science content by posing an investigable question, making a claim, forming a hypothesis, or drawing a conclusion. Requirements for students to draw on supporting evidence (e.g., observations, examples, details, illustrations, facts, data, logical reasoning) and incorporate it into their communication were apparent. Students were asked to make connections with, and refer to, big ideas in science in their communication.

Construction of knowledge about Earth science content. This rubric pertained to the extent to which the assignment facilitated students' construction of new knowledge about Earth science content and asked them to apply that knowledge to a different situation.

Indicators: Students were asked to construct new knowledge about Earth science content (e.g., concept, phenomena, idea, process) by investigating, interpreting, analyzing, synthesizing, or evaluating information. Students were also asked to apply the knowledge they gained in the assignment to a different situation.

Quality of Earth science content. This rubric pertained to the quality of the Earth science content students encountered in the assignment.

Indicators: Assignments that addressed or included high-quality Earth science content were those in which students encountered content aligned with a big idea that was significant and fundamental to the field, appropriate and challenging, and accurate.

Approach to the nature of science. This rubric pertained to the extent to which the assignment framed science as a dynamic body of knowledge developed through investigation and called for students to understand or experience how science advances through inquiry, including the tools and processes involved.

Indicators: The assignment presented or framed science as a dynamic body of knowledge developed through investigation rather than as an isolated set of facts to be memorized.

Scientific inquiry. This rubric pertained to the extent to which the assignment engaged students in all aspects of the inquiry process

Indicators: Students were asked to pose questions, select methods for use in answering those questions, carry out an investigation, and analyze and communicate results. The Appendix presents the complete rubric for this dimension.

Performance assessment quality. We also coded the quality of teachers' performance assessments by using a holistic rubric. This rubric pertained to the extent to which the performance assessment required students to apply the content they had learned and to demonstrate the understanding and skill they had gained.

Indicators: The performance assessment asked students to complete an open-ended project or solve an open-ended problem. The assessment required students to apply the content they had learned and to demonstrate the skills and/or communicate the understanding they had gained. The performance assessment provided students with clear expectations and evaluation criteria before they began the task.

Procedures

Questionnaire Administration

The research team administered postworkshop questionnaires to teachers at the end of their workshops. Teachers completed their questionnaires and submitted them to researchers directly, with a 100% response rate.

Using a commercial Web-based survey program, the research team also administered postunit questionnaires to teachers online. Teachers completed the questionnaires within 3 to 4 weeks of completing the teaching of their Earth science units. Because teachers completed the units at different times, the study team used information obtained from teachers in an earlier survey to identify each teacher's ideal window for questionnaire completion. For sixth- and eighth-grade teachers in the study, the questionnaire windows were January and February 2007. For the seventh-grade teachers, the questionnaire windows were in May. We used the online survey program to monitor response rates and followed up with teachers; we obtained a 91% response rate (n = 53).

Teacher Assignment Data Collection and Scoring

The research team collected three assignments and one performance assessment from each teacher in the study during and after his or her Earth science unit. We provided teachers in all four conditions with a list of broad criteria that would be used in coding assignments and asked them to submit assignments from three time points in their unit that they believed best represented the ideal assignments we had outlined: one assignment from Weeks 1 to 3, a second from Weeks 4 to 6, and a third from Weeks 7 to 9. In addition, we asked teachers to submit one culminating performance assessment, and we informed them, in general terms, how we would code it.

This approach meant that we did not necessarily receive typical assignments from teachers; thus, the findings from the study cannot be said to capture the full range of assignments teachers gave students. At the same time, the bias we introduced could not be expected to differentially benefit any one group because we established the criteria for coding assignments independently of the interventions being studied. Thus, the impact estimates are fair judgments of the differences among intervention groups with respect to teachers' submissions of what they saw as their best assignments and performance assessment.

A member of the study team visited each classroom to assist teachers with preparing packets for submitting assignments. The visit helped ensure that each teacher submitted a cover sheet describing the context of the assignment, as well as all relevant worksheets, instructional materials, and samples of student work that would provide coders with as comprehensive a view as possible of what the assignment entailed. For curriculum materials used as part of assignments, including IES materials, teachers were instructed to make photocopies of the relevant pages of the student or teacher guide that outlined their and their students' roles in the assignment.

Teachers across conditions submitted 197 assignments. As Table 4 shows, the assignments submitted reflected a high level of consistency with what would be expected from teachers in each condition. Among IES teachers, 42 of the 52 assignments submitted were from the curriculum. ESBD teachers included no assignments from the IES curriculum, and the distribution of types of materials submitted was similar to the Control group's distribution. Finally, the Hybrid condition teachers submitted a mix of assignments from the IES curriculum and from materials they had developed on their own.

A coding team analyzed the teacher assignments and performance assessments according to the rubrics described previously. A science education researcher with a background in teaching led the coding session. The coding team consisted of five members, all of whom were experienced middle school science teachers. The coding team members were all blind to the condition to which teachers in the study had been assigned; in addition, because teacher identifiers of all kinds were stripped from the assignments, coders were also blind to the teacher whose assignment they were coding.

The coding leader trained and calibrated the coders on each rubric independently, and the coders coded for that rubric only before moving on to the next rubric. The leader

TABLE 4
Types of Assignments Submitted by Condition

	Condition				
Assignment	*IES*	*ESBD*	*Hybrid*	*Control*	*Total*
Activities from the IES curriculum	42	0	15	0	57
Teacher-modified IES activities	5	0	6	0	11
Textbook-/publisher-developed activities	0	10	3	12	25
Activities obtained from other professional development	0	1	2	5	8
District performance task	0	1	0	7	8
Teacher-developed assignment (developed on own)	1	18	5	16	40
Teacher-developed assignment (as part of project workshop)	0	3	10	0	13
Other	0	2	2	2	6
Total	52	47	48	50	197

Note. IES = Investigating Earth Systems; ESBD = Earth Science by Design.

used actual examples and examples from other sources as training papers. For the first 25% of coding on a rubric, the coders' decisions were entered on a spreadsheet and reliability was calculated. If a discrepancy of more than 1 code point on the scale occurred, the coding leader reconvened the coders for recalibration. Overall, intercoder agreement within 1 point of the rubric scales was 87%. This level of reliability is comparable to that for other studies of teacher assignments (Shkolnik et al., 2007).

We constructed a matrix that randomly assigned the assignment and assessment artifacts to the coder so that no coder saw the same artifact twice for the same rubric, and so that each coder saw the artifacts no more than three times across the five or six rubrics. This was done to ensure that the codes for each rubric on each set of artifacts remained independent of any overall impression of the assignment that a coder might form from seeing the same assignment multiple times. Furthermore, the coders never saw an assignment twice in a row for two different rubrics. Two coders working independently scored all assignments and performance assessments.

Approach to Data Analysis

Survey Data

We fit a two-level hierarchical linear model for two of the three instructional planning variables analyzed in the study using the Stata statistical package (Version 9). Our primary interest was not the omnibus test of overall equality among all four experimental conditions, however. Rather, we were primarily interested in pairwise comparisons between each treatment group and the Control group. We were secondarily interested in pairwise comparisons among the treatment groups. In any cases for which the omnibus test was statistically significant at $p < .05$, we examined all pairwise comparisons, using a false discovery rate correction for family-wise Type I error rate (Benjamini & Hochberg, 1995). For the third instructional planning variable, overall influence on teachers' instructional planning process, we analyzed the categorical data using an ordered logistic regression model with correction for clustering of teachers within schools.

Teacher Assignments and Performance Assessments

The estimation procedures for analyzing the quality of assignments were based on procedures first developed by Newmann and colleagues (2001) in their studies conducted in Chicago Public Schools, as elaborated by study teams at the American Institutes for Research and SRI International (Shkolnik et al., 2007). The analysis first involves using many-facet Rasch measurement (MFRM; Linacre, 1989) to combine scores from individual rubrics for assignments into a single score for quality and then conducting a hierarchical linear modeling analysis to analyze treatment effects.

MFRM is a technique that statistically adjusts for the differences between both scorers and rubrics. Because the model uses data from all coders to estimate the degree of severity they exact in coding, raw scores can be adjusted for how strict or lenient a particular coder is. In addition, MFRM uses data across rubrics to adjust for the difficulty of different rubrics. For example, on some rubrics, it may be much easier to score a 4 than on others, and the model takes that difficulty into account in estimating a score for each assignment. For our analyses, we assessed the holistic rubric scores for the performance assessment quality rubric using MFRM because we had data available from two scorers and other rubrics to estimate parameters for coder severity and rubric difficulty. At the end of the analysis, the modeling procedure rescaled the raw scores from 0 to 10.

Hierarchical linear modeling is the appropriate statistical procedure for estimating impacts when data are nested, as they were in this study (Raudenbush & Bryk, 2002). In the models we tested, assignments were nested within teachers (who were part of the treatment conditions), and teachers were nested within schools. We tested separate models for the seven dependent variables: overall assignment quality, each of the five 4-point rubrics, and the overall performance assessment quality rubric. We first fit unconditional models to test the significance of variance at the assignment, teacher, and school levels. For all seven dependent variables, the unconditional variance component at the school level was nearly 0 (less than 1% of the total variance). Therefore, we adopted a two-level model (assignment within teacher) for the subsequent analyses.[2]

For each model, we included three teacher-level (Level 2) predictors. The first predictor, treatment condition, was entered as a set of three indicator variables

[2]Note that for the overall performance assessment score, with one performance assessment per teacher, we used a simpler one-level regression model.

(with the Control condition the omitted category). The models also incorporated the grade level taught (as two indicator variables for the three grade levels) and years of teaching science (as a continuous predictor). We found significant interaction effects between the treatment condition and the grade level taught and therefore incorporated those interactions in the model. The full model was as follows: Level 1 (Assignment)

$$\begin{aligned} Y_{ij} = {} & \beta_{0j} + \beta_1 IES + \beta_2 ESBD + \beta_3 Hybrid \\ & + \beta_4 G6 + \beta_5 G7 \\ & + \beta_6 IES{*}G6 + \beta_7 IES{*}G7 + \beta_8 ESBD{*}G6 + \beta_9 ESBD{*}G7 + \\ & \beta_{10} Hybrid{*}G6\ \beta_{11} Hybrid{*}G7 \\ & + \beta_{12} Years + \varepsilon_{ij} \end{aligned}$$

Level 2 (Teacher)

$$\beta_0 = \gamma_{00} + u_j,$$

where the intercept β_{0j} is modeled as a random effect of teachers, and all other predictors are fixed effects.

Corrections for multiple comparisons. Because in both the survey and hierarchical linear modeling analyses we conducted multiple comparisons within the same data set, the risk of family-wise error was increased. In particular, the likelihood of obtaining a significant result increased with multiple tests, even when the probabilities were low for detecting statistical significance. Throughout this article, we report significance levels for treatment impacts and for comparisons among the treatment groups. For both, we used a false discovery rate correction (Benjamini & Hochberg, 1995) to establish a more conservative criterion for statistical significance. Pairwise comparisons among four experimental conditions yield six statistical tests; the false discovery rate correction ensures that fewer than 5% of the reported significant results in any one family of tests are the result of a Type I error.

RESULTS

Teachers' Experiences of the Professional Development

Forms of Engagement

Table 5 shows those active learning strategies that at least half of the teachers in each condition endorsed as characteristic of their initial workshop. The teachers in each of the conditions reported different kinds of active learning strategies that largely reflected the nature of the professional development designs to which they were ex-

TABLE 5
Elements of Engagement Enacted in Workshops by Condition

IES	*ESBD*	*Hybrid*
Conducted a demonstration of a lesson, a unit, or a skill (14)	Developed a unit plan (13)	Developed a unit plan (17)
Gave a lecture or presentation (13)	Wrote reflections in a journal (11)	Practiced using student materials (16)
Practiced using student materials (13)		Wrote reflections in a journal (13)
Developed a unit plan (10)		Conducted a demonstration of a lesson, a unit, or a skill (8)
Led a small-group discussion (8)		Led a small-group discussion (8)

Note. Numbers in parentheses indicate how many teachers endorsed the item. Only items endorsed by a majority of participants as characteristic of the workshop are included in the table. IES = Investigating Earth Systems; ESBD = Earth Science by Design.

posed. For example, all of the participants in the IES condition and just more than half in the Hybrid condition reported that they conducted a demonstration of a lesson. As part of both workshops, teachers had the opportunity to demonstrate the implementation of an investigation to their colleagues. In contrast, all of the ESBD and Hybrid teachers reported that they developed a unit plan, but a few IES teachers indicated that they had not done so. In both the ESBD and Hybrid workshops, teachers said leaders emphasized unit planning. Had we asked about the nature of the plans developed in the IES workshop, we might have discovered what we would later learn from the postunit implementation survey: namely, that these were “pacing guides” rather than full unit plans. Overall, fewer different strategies were reported by more than half the participants in the ESBD condition than in the other two conditions.

Interactions With Workshop Leaders and Peers

Table 6 shows those kinds of interactions with workshop leaders and peers that at least half of the teachers in each condition identified as characteristic of their initial workshop. In contrast to teachers’ reports of active learning strategies used, the kinds of interaction reported were much more similar across condition. Nearly all teachers in each condition spent extensive time discussing implementation in different contexts: with leaders, informally with colleagues, and formally with colleagues. More than half of all teachers in each condition also reported receiving coaching and mentoring as part of their professional development. The only clear distinction among conditions was that compared with teachers in the IES condition, a higher percentage of teachers in the ESBD and Hybrid conditions reported receiving feedback on lesson plans they had developed. This contrast is not surprising and shows consistency with the professional development models, given that the plan was for teachers in both the ESBD and Hybrid conditions to develop, receive feedback on, and revise unit plans.

TABLE 6
Interactions With Workshop Leaders and Peers by Condition

IES	*ESBD*	*Hybrid*
Communicated with the leaders of the activity concerning implementation (13)	Developed curricula or lesson plans that other participants or the activity leaders reviewed (12)	Developed curricula or lesson plans that other participants or the activity leaders reviewed (17)
Met informally with other participants to discuss implementation of my unit (12)	Communicated with the leaders of the activity concerning implementation (12)	Communicated with the leaders of the activity concerning implementation (17)
Practiced under simulated conditions, with feedback (12)	Met formally with other participants to discuss implementation of my unit (12)	Met informally with other participants to discuss implementation of my unit (16)
Met formally with other participants to discuss implementation of my unit (11)	Met informally with other participants to discuss implementation of my unit (11)	Met formally with other participants to discuss implementation of my unit (15)
Received coaching or mentoring (9)	Received coaching or mentoring (9)	Received coaching or mentoring (9)
Developed curricula or lesson plans that other participants or the activity leaders reviewed (9)		

Note. Numbers in parentheses indicate how many teachers endorsed the item. Only items endorsed by a majority of participants as characteristic of the workshop are included in the table. IES = Investigating Earth Systems; ESBD = Earth Science by Design.

Judgments of Coherence

Coherence ratings (i.e., ratings of how consistent the workshops goals were with district goals for students) were very high for all of the conditions. Teachers in all three conditions rated the workshops between 11.7 and 13.4 on a scale from 1 to 14, with 14 representing the highest possible coherence. Ratings were highest for the Hybrid condition and lowest for the IES condition, but the differences between the two sets of ratings were not statistically significant (see Figure 1).

Preparation for Inquiry

Teachers assigned to the two conditions that promoted inquiry-based approaches to teaching science rated their professional development higher in terms of how well their workshops prepared them for inquiry teaching than did teachers in the other condition. Teachers in the IES and Hybrid conditions rated the workshops at 2.1 on a scale from 0 to 3, with 3 representing the highest possible preparation for inquiry. These ratings were higher than the ratings of teachers in the ESBD condition, which did not explicitly promote inquiry-based teaching as part of its model (see Figure 2).

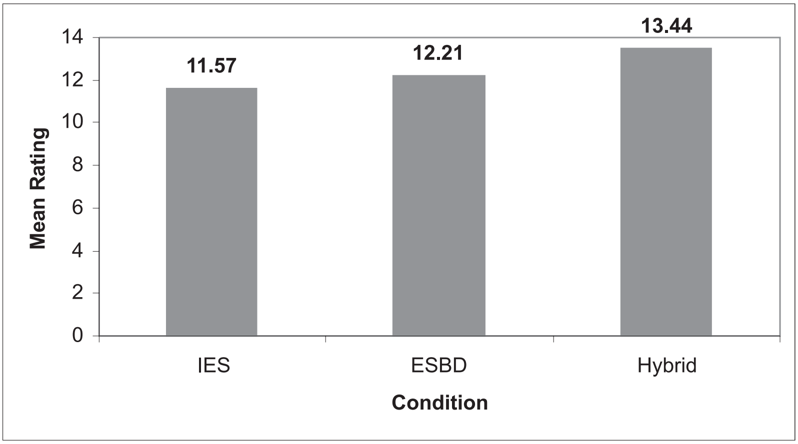

FIGURE 1 Ratings of the coherence of the workshops. IES = Investigating Earth Systems; ESBD = Earth Science by Design

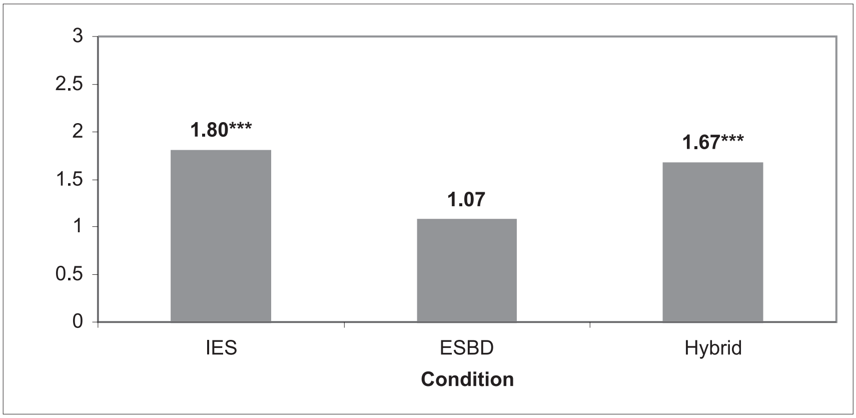

FIGURE 2 Ratings of relative emphasis on science inquiry in the workshop. IES = Investigating Earth Systems; ESBD = Earth Science by Design. ***Pairwise comparisons to ESBD significant at $p < .0001$.

Reported Impacts on Instructional Planning

There were statistically significant differences among the four conditions with respect to teachers' reports of the project's influence on how they planned instruction in science. Table 7 shows the distribution of responses to the implementation questionnaire item related to overall impacts on planning. All of the ESBD and

TABLE 7
Reported Influences of Project on Instructional Planning in Science by Condition

Influence	*IES*	*ESBD*	*Hybrid*	*Control*
Not at all	3	0	0	5
Somewhat	7	2	5	6
Very much	3	9	8	2

Note. Model $F(3, 16) = 8.23$, $p < .01$. IES = Investigating Earth Systems; ESBD = Earth Science by Design.

Hybrid teachers reported changes in their practice; in contrast, at least some teachers in the IES and Control conditions reported no changes.

The amount of change in instructional planning reported was consistent with the program designs, and there was also an overall effect of condition on reported changes to planning, $F(3, 16) = 8.23$, $p < .01$. Post hoc tests revealed that the ESBD and Hybrid intervention conditions outscored both the IES and Control conditions.

With respect to the use of instructional materials, there were significant overall effects of condition for all materials queried in the implementation survey; for the most part, those effects reflected the designs of the different professional development programs (see Table 8). Post hoc pairwise comparisons indicated that teachers in the IES condition reported making significantly less use of the district textbook than teachers in all other conditions, consistent with the instructions we gave them. Teachers in each of the intervention conditions reported making significantly more use of Internet-based resources, such that IES teachers could find materials that supplemented activities in their student guides and ESBD and Hybrid teachers could find visualizations. Teachers in the ESBD and Hybrid conditions reported making more use of visualizations, as they had been instructed to do, than did teachers in the IES and Control conditions. Teachers in the ESBD condition also made more use of materials they themselves had created than did teachers in the Hybrid condition, and teachers in the Hybrid condition, in turn, said they made more use of self-created materials than did teachers in the IES condition. Finally, teachers in the Hybrid condition reported making more use of materials colleagues had developed than did teachers in the IES and Control conditions. Because these last findings were unexpected, we address them at greater length in the Discussion. All statistically significant pairwise comparisons within each type of instructional material were controlled for a false discovery rate of $p < .05$.

Impacts on Assignment Quality

Figure 3 shows the distribution of standardized assignment quality scores by condition for each dimension of the rubric and for overall assignment quality. Each ru-

TABLE 8
Changes in Instructional Materials Used in Earth Science Units

Type of Material	*Condition*	*M*	*SD*	*Omnibus* χ^2
District-selected textbook	IES	1.15	0.38	42.95***
	ESBD	3.18	1.60	
	Hybrid	3.23	0.93	
	Control	3.38	0.87	
Internet	IES	5.08	0.76	23.32***
	ESBD	5.64	0.50	
	Hybrid	4.92	0.76	
	Control	3.62	1.76	
Visualizations	IES	4.31	0.85	33.89***
	ESBD	5.73	0.47	
	Hybrid	5.31	0.85	
	Control	4.54	0.77	
Materials I created	IES	2.85	1.72	27.23***
	ESBD	5.18	0.87	
	Hybrid	3.85	0.99	
	Control	4.46	0.66	
Materials colleagues created	IES	1.77	1.30	15.58**
	ESBD	2.82	1.60	
	Hybrid	3.77	1.30	
	Control	2.15	1.41	

Note. A 6-point scale was used (1 = *did not use*, 2 = *used much less*, 3 = *used a little less*, 4 = *used about the same*, 5 = *used a little more*, 6 = *used a lot more*). IES = Investigating Earth Systems; ESBD = Earth Science by Design.

$^{**}p < .01$. $^{***}p < .001$.

bric score was standardized to a mean of 0 and standard deviation of 1 across the entire sample. The box-and-whisker plot allows one to see, for example, that the nature of science ratings for both the ESBD and Control conditions were substantially lower than the overall mean of 0, and their dispersion was attenuated relative to the other two conditions as well.

With respect to quality, teachers in the IES and Hybrid conditions, who had access to curriculum materials written by experts, submitted assignments whose overall quality the coders rated higher than that of assignments submitted by teachers in the Control condition. Furthermore, among the treatment conditions, assignment quality was significantly higher in the IES and Hybrid conditions than in the ESBD condition (see Table 9). At the same time, the impact of assignments did not provide evidence that any of the programs had affected teachers' assessment practices.

Within the specific rubric dimensions, many of the effects may be attributable to substantial differences in how assignments were rated with respect to their emphasis on scientific communication and on inquiry. Rubric scores for three dimen-

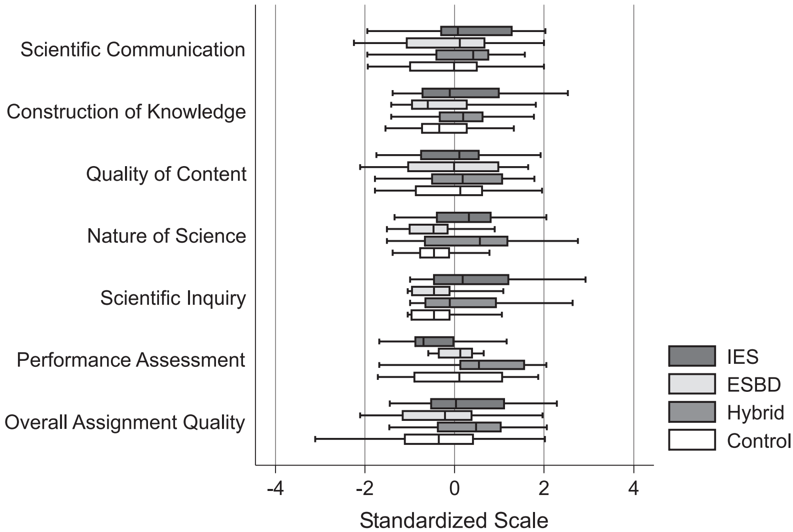

FIGURE 3 Distribution of scale scores by condition for assignment quality (combining grade levels). IES = Investigating Earth Systems; ESBD = Earth Science by Design.

sions—scientific communication, nature of science, and scientific inquiry—were higher for assignments submitted by teachers in the IES and Hybrid conditions than for assignments submitted by Control group teachers. It is likely that these differences partly reflect what writers of the IES curriculum intended for the investigations at the heart of the curriculum. The investigations require students to pose questions, collect and analyze data, and communicate results of their investigations. If a teacher assigned one of these investigations from start to finish, that assignment would have rated high on the rubric developed for the study.

Even when students are not engaged in conducting actual investigations, several of the IES activities engage students directly in thinking about the nature of science. In the Astronomy module, for example, in an investigation focused on helping students understand the Big Bang theory, students read that it has been problematic for scientists to collect evidence to support particular theories of the origins of the universe. Students also learn about how advances in instrumentation enabled scientists to collect better evidence for supporting theories; furthermore, students are asked to reflect on what evidence is missing for the Big Bang theory and to consider more generally what the main problems are in studying objects in the universe.

The differences between IES-based assignments and those submitted by Control group teachers are not entirely explained by the strength of the IES materials

TABLE 9
Effect Sizes for Ratings of Assignment Quality

Condition	*ESBD*	*Hybrid*	*Control*
Overall assignment quality			
IES	1.10*	0.15	0.85*
ESBD	—	0.95*	0.25
Hybrid		—	0.70*
Overall performance assessment quality			
IES	0.28	0.86	0.41
ESBD	—	0.58	0.13
Hybrid		—	0.45
Scientific communication			
IES	0.78*	0.21	0.60*
ESBD	—	0.56*	0.18
Hybrid		—	0.38
Construction of knowledge			
IES	0.85*	0.19	0.61*
ESBD	—	0.66*	0.24
Hybrid		—	0.42
Quality of Earth science content			
IES	0.23	0.26	0.24
ESBD	—	0.49	0.02
Hybrid		—	0.51
Nature of science			
IES	1.28*	0.10	1.14*
ESBD	—	1.18*	0.15
Hybrid		—	1.04*
Scientific inquiry			
IES	0.91*	0.36	0.77*
ESBD	—	0.55	0.14
Hybrid		—	0.41

Note. IES = Investigating Earth Systems; ESBD = Earth Science by Design.
*Statistically significant difference controlling for false discovery rate of .05 among six comparisons.

alone; they are also partly a function of the lower quality ratings assigned to the many teacher-developed materials submitted by Control group teachers. One Control group teacher's assignment on planets and their satellites, for example, asked students to conduct research on the inner planets in the Solar System and present to the class information they found about their planet's name, distance from the Sun, diameter, revolution and rotation period, and number of moons. The teacher assessed the students on the accuracy of the information, all of which students could look up using either the textbook or the Internet. The fact that the assignment did not engage students in considering how scientists develop knowledge about planets meant that it scored low on the nature of science rubric. For another astronomy assignment provided by another Control group teacher, students had to demonstrate their understanding of moon phases by drawing the phases on a sheet of paper. The assignment did not engage students with any aspect of how phases are observed or about how they might be different if viewed from a different perspective (e.g., from outer space, as opposed to from the Earth), thus providing limited opportunity for students to reflect on how scientists know about the phases of the moon. The rubric for the task, moreover, indicated that students were graded not for the accuracy of their drawings but for creativity, neatness, and proper coloring and labeling. Of these elements, only labeling could be said to be directly related to knowledge of the phases of the moon, and it is unclear whether this assignment calls for development of deep understanding of the moon phases at all.

DISCUSSION

In this section, to trace a path from design to professional development experience to teachers' designs for instruction, we draw connections between findings from different sources and phases of enactment of the professional development designs. In most respects, the pattern of results across the survey and teacher assignment analyses coheres and points to important ways that each professional development program impacts (or fails to impact) teachers' design of instructional experience for students. For example, congruent with the focus of the ESBD or the Design program of professional development, nearly all teachers in this group reported that they developed unit plans as part of their workshops. They also reported making extensive use of visualizations—an important component of the Design program—in developing lessons. When submitting assignments, they submitted a mixture of teacher-created and publisher-created lessons, a similar mixture to that for the Control group (who got no professional development targeting the design of instructional experiences for students). This mixture can be readily interpreted in light of the guidance that the Design program gave to teachers about using whatever materials were available in constructing units (except the inquiry-based curriculum, to which they had no access). Congruent with the design of the IES or the Adoption program of professional development, teachers in this condition reported that lessons were part of the inquiry-based curriculum, design, and demonstrations in their 2-week workshop. In planning lessons, teachers in this condition reported making little to no use of the district technology or materials that they or their colleagues developed. They used the Internet, as instructed, as a supplementary resource, and they submitted lessons principally from the curriculum as part of the assignment study. Finally, congruent with the design of the Hybrid or Principled Adaptation condition, teachers in this condition reported that their key professional development experiences consisted of developing a unit plan and engaging in investigations that were part of the inquiry-oriented curriculum. As intended by the program, teachers reported that their instructional planning process changed as a result of participating in the program and that they made more use of visualizations—another part of the design approach incorporated in the Principled Adaptation program. They submitted a mixture of lessons from the inquiry-oriented curriculum and teacher-made materials, as might be expected given the program leaders' guidance to use at least half of the investigations in each curriculum module.

The overall congruence between the ways in which the professional development programs were enacted and experienced by teachers probably explains the large impacts on the assignment quality of those programs that provided teachers with high-quality curriculum materials. As intended, teachers from both the Adoption and Principled Adaptation programs submitted assignments that used the materials the programs had prepared them to use. The overall effects on assignment quality of these programs reflect a trend we observed in the data for assignments: Selections from the IES curriculum were rated higher than any other type of material submitted by teachers. At the same time, the ratings of assignment quality were significantly higher among the Principled Adaptation teachers than among the Control teachers, even though Principled Adaptation teachers submitted many assignments that they or their colleagues developed. This finding suggests that allowing teachers to choose what to enact did not significantly diminish the depth of student opportunities to learn Earth science afforded by the inquiry-oriented science curriculum.

In many respects, our findings are consistent with Cohen and Hill's (2001) findings regarding effective professional development in mathematics curricula, with three important exceptions. The findings corroborate their finding that only teachers who received professional development in which they were provided materials and strategies that embodied the principles of reform made significant changes to their practice consistent with those principles, whereas teachers who received either generic strategy or no professional development tended to make few changes to their practice. Our findings are in a different domain from theirs (science) and so extend their insights about the importance of materials for practice in professional development for promoting change. Furthermore, in contrast to their study, we used a random assignment design, reducing bias that can be associated with teacher selection into professional development opportunities. Finally, Cohen and Hill were not concerned in their study with comparing different approaches to helping teachers design instructional experiences; rather, theirs was a study focused on illuminating learning problems associated with implementing reforms at scale in a large state.

With respect to our study's focus on the design of instructional experiences, our findings suggest ways in which some frameworks recommending that teachers play more significant roles in designing and adapting curriculum in professional development may be underspecified. The Understanding by Design framework

(Wiggins & McTighe, 1998), which inspired the development of the Design program in this study, does not specify which curriculum materials teachers should choose; rather, it specifies a process for selecting materials based principally on goals for student learning and on pedagogical strategies the materials will allow teachers to enact with students. Connelly and Clandin (1988) gave close attention to curriculum inquiry in their treatment of teachers as curriculum designers, but they argued that curriculum materials may have potential beyond what is evident from the quality of the materials themselves that can be brought out by good teaching. Indeed, materials may be "more than the sum of their parts," and our study did not examine enactment of the materials per se, but we did find that the materials that teachers chose to send us mattered when it came to lesson quality. When teachers had access to materials that experts had designed for the purpose of promoting inquiry and exposing students to the nature of science, and when their professional development encouraged them to use those materials in teaching, they were more likely to submit lessons that were judged to be of higher overall quality than were teachers who did not get professional development. In contrast, simply learning how to select materials based on goals and pedagogical strategies was not sufficient to provide teachers with skills needed to select higher quality materials.

In their study of a model for developing preservice teachers' skill in designing inquiry-based lessons, Schwarz and Gwekwerere (2007) were able to document benefits to using a framework that did not include providing specific curriculum materials for teachers to use. The framework they presented to teachers focused on eliciting student ideas, constructing student-led investigations of phenomena, and helping students build scientific models and apply those models to help them understand new phenomena. A number of teachers shifted from designing lessons that were primarily activity focused or didactic to designing lessons that were judged to be more inquiry based from the start to end of the lesson. One explanation for the pattern of results that, on the surface, appears different from our own results is that their framework was more closely aligned with the framework being used to evaluate lessons. Understanding by Design as a framework was designed to be cross-curricular; the pedagogical strategies emphasized are not tied to learning within a subject matter discipline.

Other researchers have pursued the direction of more closely aligning frameworks for design with processes associated with subject matter learning. For example, mapping models for the design of instructional materials to subject matter processes and content is the focus of Edelson's (2001) Learning-for-Use framework. That framework maps onto processes associated both with science inquiry and with principles of constructivist learning more broadly. In the article that describes this framework, Edelson presented an example that he and his team developed to reflect the framework for teaching Earth science. He suggested that it is an "open issue" whether teachers can use the framework, however, because he did not specify criteria for determining whether curricula reflect the principles of the framework, and he did not have a contrasting framework to compare his approach against.

Comparing Schwarz and Gwekwerere's (2007) and Edelson's (2001) studies to our own findings suggests some ways in which our results answer Edelson's call for comparative analysis of different approaches to teaching teachers how to design instructional materials. Neither of these earlier studies included a comparison group that either received no program or class or received some alternative approach to helping teachers design instructional experiences. Schwarz and Gwekwerere suggested that their own class might have been improved had they incorporated more curriculum review and analysis activities. They speculated that, compared with the class they taught, students in such a class might have learned more. This class would have been similar to our own Principled Adaptation program; our findings provide some evidence for their conjecture, at least for in-service teachers similar to those in our study. A possible parallel comparison study Edelson might have conducted would have compared an approach that prepares educators to use the Learning-to-Use framework with and without his own curriculum materials for designing instructional experiences for students.

We do hypothesize that the approach of more closely aligning the frameworks taught in the Design program with processes and content of science learning might be a useful strategy for enhancing the effectiveness of that program. Schwarz and Gwekwerere's (2007) success in changing preservice teachers' planning processes and conceptions of inquiry-based learning suggest to us that if the Design program incorporated more explicit attention to scaffolding the process of student inquiry and reasoning with scientific models, teachers might be able to design instructional materials of higher quality than would teachers who do not receive any professional development at all in the design of instructional materials.

It is still an open question to us whether, without examples of high-quality curriculum materials, the lessons teachers would design or submit to researchers as part of a study such as ours would be of as high a quality as lessons submitted by teachers in the Adoption or Principled Adaptation programs of professional development that were examined in the current study. We acknowledge that the fact that IES materials used by teachers in the Adoption and Principled Adaptation programs were designed in such a way that scores on the independent rubric could be expected to be higher than scores for materials that the teachers themselves developed does not diminish the significance of our findings. We conjecture that this is in part because professional curriculum developers engage in an iterative process of development, such that different ideas are implemented in materials, tested in the classroom, and then revised. Although teachers certainly engage in curriculum revision, and curriculum revision is one professional development strategy that has been promoted by scholars (e.g., Singer, Krajcik, Marx, & Clay-Chambers, 2000), teachers' approach to revision is different from that of researchers and curriculum developers. A number of projects that have engaged teachers in the codesign of materials have found that it takes a considerable amount of time for teachers to begin to "think like designers" and apply principles of how people learn to designing curriculum materials for students (Penuel, Roschelle, & Shechtman, 2007; Reiser et al., 2000). To note this fact is not to say that teachers cannot engage in design; in fact, many curriculum developers are former teachers. We do, however, see that the practice one gains from being a member of a professional curriculum design community and becoming disciplined by those practices to see curriculum potentials in a particular way prepares individuals in a qualitatively different way to design instructional materials than do preservice or in-service programs for teachers that are of a limited duration. To prepare teachers to engage in those practices in ways similar to professional curriculum designers would likely take much more time than is practical under the current division of labor in U.S. schools.

Some of the findings that disappointed us are more similar to past findings regarding professional development focused on improving the quality of classroom assessments. Our study did not find evidence that any of the three programs was effective in improving the quality of performance assessments for students at the end of their Earth science units, despite the fact that the Understanding by Design model underlying all three programs places emphasis on performance assessment and that model underlay all three programs of professional development. Even in the Adoption condition, for which teachers could submit assessments from the curriculum, performance assessment scores were no higher than in the Control condition. This finding may be attributable to a mismatch between IES's embedded assessments and the rubric, which was based on Wiggins and McTighe's (1998) definitions of assessment quality, but it may also point to the challenge of providing professional development that significantly influences teachers' assessment design skills. But other researchers have found it similarly challenging to improve assessment design skills, because assessment design requires not only deep content knowledge and knowledge for teaching but also knowledge of properties of valid, fair, and reliable tests (Hunt & Pellegrino, 2002; McMunn, McColskey, & Butler, 2004; Ruiz-Primo & Furtak, 2007; Shepard, 1997; Yarnall, Shechtman, & Penuel, 2006).

At least one project (Shepard, 1997) that was successful in changing teachers' ideas about and practices of performance assessments found that change was conditional on a number of factors. In this particular project, teachers worked with experts in assessment over the course of multiple sessions to design, test, and discuss performance assessments. The researchers studying the project reported modest changes in teachers' assessment practices but concluded that significant change could be achieved from such professional development efforts only when there was significant discussion of the assessments and when leaders challenged teachers' ideas that were incompatible with the design principles being espoused on the project (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997). In our project, there was limited interaction between professional development providers and teachers over the course of the year with respect to assessment design strategies, so the kinds of conditions Borko and colleagues (1997) believed were critical were not characteristic of our particular study setting.

Limitations of the Study

Despite the strengths of our overall research design, not all threats to validity were eliminated by random assignment. For example, differences in the skill and approach of the different staff leading the professional development workshops may have accounted for differences in outcomes. In addition, the teachers in each group ended up working collaboratively with one another as part of their workshops. That collaborative work may have resulted in convergence within their program's group, as teachers sought to align their unit designs with one another. This may have been particularly the case in the Hybrid condition, in which we observed the greatest degree of collaboration. In one respect, variations in enactment of designs are endogenous experimental factors and thus not a threat to validity. It is hard to imagine, however, that teachers' work with one another did not somehow affect the results of the study, even though we required individual teachers to submit their own assignments for the purposes of our analysis. Past studies, including those by researchers on our team, have shown significant effects from collegial help on patterns of implementation of reform programs (Penuel, Frank, & Krause, 2006). Furthermore, other research has suggested that intentional efforts to promote collaboration can be a strength of professional development designs (Garet et al., 2001; Penuel, Fishman et al., 2007) rather than a contaminant in a research study, a finding with which study team members were familiar.[3]

Our self-report measures of instructional planning are subject to bias, which is another potential limitation of the study. The ESBD program, in which design was emphasized, and the Hybrid program, in which planned adaptation was emphasized, prepared teachers to know that their planning should differ from what they did before participating. Accordingly, a question about how their planning changed was likely to yield positive effects. Alternative methods, such as think-aloud protocols or direct observations of teachers' planning, might yield more valid results; however, it might be difficult for researchers to conduct such studies without creating a contrived situation that itself introduces bias in the measurement of instructional planning.

[3]Although we recognized the need to discourage collaboration as much as possible to preserve the integrity of the study design, we also felt ambivalent about our decision. This ambivalence speaks to one challenge that can arise when implementing random assignment designs when protecting against threats to validity interferes with what professional developers believe is best practice.

Another study limitation was the absence of a measure of prior instructional quality; some variance may have been due to initial differences in teaching quality among teachers in the study. Our study relied on a relatively small sample; thus, even with random assignment, prior differences in instructional quality could have shaped the outcomes measured here. Future studies might use a longitudinal design with a baseline year after initial recruitment to collect data on prior instruction; that solution would provide a direct prior measure, but the risk of attrition would increase.

We also note that our assignments represented a sample of those that students encountered in their Earth science units and that the sample was no doubt biased in that teachers submitted the most challenging and interesting assignments to our study team. Teachers, not researchers, selected the assignments to be given, and the selection process was not random. Moreover, the assignments they chose may have focused on one or two aspects of quality on which raters based their judgments. Our guidance to teachers, to be sure, was the same across conditions; however, insofar as professional development activities may have increased teachers' sensitivity to the quality dimensions on which their assignments were ultimately scored, some interaction of treatment condition may have influenced their choice of the assignments submitted.

We recognize that measuring the impact of professional development programs like those that were the focus of our study requires multiple methods to corroborate findings. Other multimethod studies of teacher practice (e.g., Camburn & Barnes, 2004) have found that incorporation of multiple measures of instruction does not always lead to converging findings. Different measures of instruction yield different kinds of insights into potential impacts on teaching practice, particularly with respect to the design work teachers do in the midst of teaching to adapt prepared lessons to student interests and needs. Understanding this latter form of adaptation is critical because it represents an important instruction design arena that teachers enter every day.

Because our study also incorporated classroom observations, we can compare and contrast these study results with those of our observational study, details of which can be found in Penuel, Benbow, and colleagues (2009). In that study, we found that students of teachers in all three programs had a better understanding of the connection between their current classroom assignment and the big idea the teacher was teaching. We also found that compared with teachers in the Control condition, teachers in the Design and Adaptation programs were more likely to engage students in assignments that required them to interpret texts or data in class. We found no impacts on the classroom assessment practices of teachers in any of the professional development programs, however. This pattern of results differed slightly from our study results for the assignment analysis: All three programs showed some benefit relative to the Control condition. At the same time, our finding that none of the programs had an effect on assessment was consistent with the results of this study.

Directions for Future Research

We believe that Bruner's (1960) assertion that it is "too much" to ask of teachers to design instructional experiences for students is in at least one important sense wrong. Our study has found that teachers can, in fact, be asked to adapt curricula to their local context, because programs can be designed that demonstrate that the adaptation of curricula to develop lessons for students does not diminish the quality of the curricular materials themselves. Furthermore, our study findings indicate that programs that help teachers learn principles for adapting curricula yield lessons that are of higher quality than those taught by teachers who receive no professional development but find themselves, as all teachers do, having to adapt materials to their local contexts.

Put another way, this study provides evidence that designing professional development to support what Remillard (2005) called a "participating with" concept of curriculum use can positively impact teaching practice. This concept views teachers and their materials as influencing one another: Just as teachers can change materials by using them, using those materials can change teachers. Although Remillard (2005) herself used this concept to describe different approaches to describing actual curriculum use, we have extended it to apply to professional development designs and have found it useful as a way to characterize differences in those designs. Furthermore, although we have not presented evidence that curriculum use changes teachers' self-concepts, as some researchers have sought to do, we have provided evidence of the potential of active participation with curriculum materials to influence what teachers do.

Many policymakers and curriculum writers through the years have expressed frustration when they have come to realize that teachers' adaptations of curricula are unavoidable. But from our study, we conclude that teachers' adaptations do not necessarily diminish the quality of student opportunities to learn and that programs can make a difference in ensuring that this is the case. At the same time, simply delivering books, materials, and science kits to teachers without significant professional development is unlikely to be sufficient to promote curriculum implementation. We believe that our study findings show that curriculum design and professional development design should go hand in hand: Writers of curricula need to consider the ways in which teachers can adapt their materials, and they need to develop plans for professional development that prepares teachers to make adaptations that are congruent with the purposes of their written curricula.

Such a conclusion will hardly be surprising to researchers, policymakers, and curriculum writers. All have observed the troubles teachers face in enacting ambitious curricula that reflect learning sciences principles. At the same time, in public debates about how to improve science instruction, the idea that materials can be "teacher-proofed" resurfaces again and again (Atkin & Black, 2003). Our study provides important evidence to support arguments for making professional development that prepares teachers to adopt and adapt materials into an essential component of plans to scale new curricula.

At the same time, we found that these programs were less successful than intended in helping teachers design classroom assessments to measure what students had learned from a unit's activities. The promise of formative assessment is great (Black & Wiliam, 1998), but so too are the challenges of preparing teachers to design and enact assessments that are likely to improve student learning. Other studies of projects intended to improve teachers' classroom assessment practice have tended to show only modest gains, if they show gains at all (Ruiz-Primo & Furtak, 2007; Shepard, 1997; Yarnall et al., 2006). This study's finding that there was no evidence of impact on assessment practice is consistent with the results of earlier studies, which points to the need for further research and development in this area.

Finally, despite their limitations, we conclude that teacher assignments provide valuable insights into teachers' assumptions about what students should learn and how they should learn it—a critical component of teachers' pedagogical design capacity. We agree with the critique of these methods in earlier analyses of teacher assignments that rubric scores are not reliable enough to warrant making judgments about individual teacher performance. But our study results indicate that, even when measured with error and taking into account the fact that individual teachers' instruction can be highly variable, significant program impacts can be found using this approach. This finding is important because researchers need multiple tools to measure instruction, each of which can provide a different perspective on the complex acts of teaching science in schools.

We hope that other learning scientists concerned with how best to bring their curricular innovations to scale will conduct research studies to investigate how best to influence teachers' pedagogical design capacity. Our study has examined only one aspect of this capacity: how teachers plan and organize student opportunities to learn. Other researchers have focused on the need to examine how teachers learn from materials themselves and are transformed by those materials, and we agree that these are fruitful avenues for learning sciences researchers to pursue. We as a field also need a better understanding of how to prepare teachers to adapt instruction in light of what students know and can do, as some learning sciences researchers have also investigated. The more we understand about how best to promote teachers' learning about how to plan and coordinate instruction with high-quality materials, the more likely it will be that our efforts to design those materials achieve a broader impact.

ACKNOWLEDGMENTS

This work was supported by U.S. Department of Education Grant R305M050226. Any opinions, findings, and conclusions or recommendations expressed in this material are our own and do not necessarily reflect the views of the U.S. Department of Education. We wish to thank those who managed and coordinated data collection on this project, including Christine Korbak, Angeline Reyes, Amy Hafter, and Michelle Phillips. We also wish to thank Barry Fishman, Harold McWilliams, Carla McAuliffe, Colin Mably, Ann Benbow, and Margaret Hayden for their comments on earlier drafts of this manuscript, and Iris Tabak and three anonymous reviewers for their comments on the manuscript.

REFERENCES

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

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 26*(9), 6–8, 14.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57,* 289–300.

Ben-Peretz, M. (1990). *The teacher-curriculum encounter: Freeing teachers from the tyranny of texts.* Albany: State University of New York Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education, 13,* 259–278.

Borko, H., Stecher, B. M., Martinez, F., Kuffner, K. L., Barnes, D., Arnold, S. C., et al. (2006). *Using classroom artifacts to measure instructional practice in middle school science: A two-state field test.* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Brown, A. L., & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Hillsdale, NJ: Erlbaum.

Brown, M. W. (2002). *Teaching by design: Understanding the intersection between teacher practice and the design of curricular innovations.* Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Brown, M. W. (2008). Toward a theory of curriculum design and use: Understanding the teacher-tool relationship. In J. T. Remillard, G. Herbel-Eisenman, & G. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17–36). New York: Routledge.

Brown, M. W., & Edelson, D. (2003). *Teaching as design: Can we better understand the ways teachers use materials so we can better design materials to support their changes in practice.* Evanston, IL: Center for Learning Technologies in Urban Schools.

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

Bullough, R. (1992). Beginning teacher curriculum decision making, personal teaching metaphors, and teacher education. *Teaching and Teacher Education, 8,* 239–252.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices.* Portsmouth, NH: Heinemann.

Camburn, E., & Barnes, C. (2004). Assessing the validity of a language arts instruction log through triangulation. *Elementary School Journal, 105,* 49–74.

Carlsen, W. S. (1991). Effects of new biology teachers' subject-matter knowledge on curricular planning. *Science Education, 75,* 631–647.

Carlsen, W. S. (1993). Teacher knowledge and discourse control: Quantitative evidence from novice biology teachers' classroom. *Journal of Research in Science Teaching, 30,* 417–481.

Clare, L., & Aschbacher, P. R. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment, 7*(1), 39–59.

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

Collopy, R. (2003). Curriculum materials as a professional development tool: How a mathematics textbook affected two teachers' learning. *Elementary School Journal, 103,* 227–311.

Connelly, F. M., & Ben-Peretz, M. (1980). Teachers' roles in the using and doing of research and curriculum. *Journal of Curriculum Studies, 12*(2), 95–107.

Connelly, F. M., & Clandin, D. J. (1988). *Teachers as curriculum planners: Narratives of experience.* Toronto and New York: OISE Press and Teachers College Press.

Crawford, B. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching, 37,* 916–937.

Crawford, V., Korbak, C., & Gong, B. (2003). *Earth Science by Design evaluation: Summer institute evaluation report.* Menlo Park, CA: SRI International.

Davis, E. A. (2008, March). *Beginning elementary teachers' ideas about inquiry and effective science teaching: A longitudinal study.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

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

Davis, E. A., & Varma, K. (2008). Supporting teachers in productive adaptation. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Designing coherent science education* (pp. 94–122). New York: Teachers College Press.

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.

Drake, C., & Sherin, M. G. (2006). Practicing change: Curriculum adaptation and teacher narrative in the context of mathematics education reform. *Curriculum Inquiry, 36*(2), 153–187.

Edelson, D. C. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching, 38,* 355–385.

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,* 643–658.

Fishman, B. J., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *Journal of the Learning Sciences, 13,* 43–76.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. *Prevention Science, 6*(3), 151–175.

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,* 915–945.

Glascow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research into practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93,* 1261–1267.

Hunt, E., & Pellegrino, J. W. (2002). Issues, examples, and challenges in formative assessment. *New Directions for Teaching and Learning, 89,* 73–85.

Ingersoll, R. M. (2003). *Who controls teachers' work? Power and accountability in America's schools.* Cambridge, MA: Harvard University Press.

Johnson, C. C. (2007). Whole-school collaborative sustained professional development and science teacher change: Signs of progress. *Journal of Science Teacher Education, 18,* 629–661.

Judson, E., & Lawson, A. E. (2007). What is the role of constructivist teachers within faculty communication networks? *Journal of Research in Science Teaching, 44,* 490–505.

Kali, Y., Linn, M. C., & Roseman, J. E. (2008). *Designing coherent science education: Implications for curriculum, instruction, and policy.* New York: Teachers College Press.

Keys, C. W., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching, 38,* 631–645.

Krajcik, J., & Reiser, B. J. (2004). *IQWST: Investigating and Questioning our World Through Science and Technology.* Ann Arbor, MI: University of Michigan.

Lang, M., Drake, S., & Olson, J. (2006). Discourse and the new didactics of scientific literacy. *Journal of Curriculum Studies, 38*(2), 177–188.

Li, J., Klahr, D., & Siler, S. (2006). What lies beneath the science achievement gap: The challenges of aligning science instruction with standards and tests. *Science Educator, 15*(1), 1–12.

Lin, H.-T. (2008). *Exploring the role of showing design intent in supporting curriculum modifications.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Lin, H.-T., & Fishman, B. J. (2004). Supporting the scaling of innovations: Guiding teacher adaptation of materials by making implicit structures explicit. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 617–618). Santa Monica, CA: Erlbaum.

Linacre, J. M. (1989). *Many-facet Rasch measurement.* Chicago: MESA Press.

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.

Marton, F., Tsui, A. B. M., Chik, P. P. M., Ko, P. Y., Lo, M. L., & Mok, A. I. C. (2004). *Classroom discourse and the space of learning.* Mahwah, NJ: Erlbaum.

Marx, R. W., Blumenfeld, P., Krajcik, J. S., & Soloway, E. (1998). New technologies for teacher professional development. *Teaching and Teacher Education, 14,* 33–52.

Marx, R. W., & Harris, C. J. (2006). No Child Left Behind and science education: Opportunities, challenges, and risks. *Elementary School Journal, 106,* 467–477.

Matsumura, L. C., Patthey-Chavez, G. G., Valdez, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *Elementary School Journal, 103,* 3–25.

McCutcheon, G. (1980). How do elementary teachers plan? The nature of planning and influences on it. *Elementary School Journal, 81,* 4–23.

McLaughlin, M. W. (1976). Implementation as mutual adaptation: Change in classroom organization. *Teachers College Record, 77,* 339–351.

McMunn, N., McColskey, W., & Butler, S. (2004). Building teacher capacity in classroom assessment to improve student learning. *International Journal of Educational Policy, Research, & Practice, 4*(4), 25–48.

Meyer, H.-D., & Rowan, B. (2006). Institutional analysis and the study of education. In H.-D. Meyer & B. Rowan (Eds.), *The new institutionalism in education* (pp. 1–13). Albany: State University of New York Press.

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Newman, F. M., Lopez, E., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report.* Chicago, IL: Consortium on Chicago School Research.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84.

Parke, H. M., & Coble, C. R. (1997). Teachers designing curriculum as professional development: A model for transformational science teaching. *Journal of Research in Science Teaching, 34,* 773–789.

Pea, R. D., & Collins, A. (2008). Learning how to do science education: Four waves of reform. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Designing coherent science education* (pp. 3–12). New York: Teachers College Press.

Penuel, W. R., Benbow, A., Mably, C., McWilliams, H., McAuliffe, C., & Hayden, M. M. (2009). Teaching for understanding in Earth science: Comparing impacts on planning and instruction in three professional development designs for middle school science teachers. *Journal of Science Teacher Education, 20*(5), 415–436.

Penuel, W. R., Fishman, B. J., Gallagher, L. P., Korbak, C., & Lopez-Prado, B. (2009). Is alignment enough? Investigating the effects of state policies and professional development on science curriculum implementation. *Science Education, 93*(4), 656–677.

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,* 921–958.

Penuel, W. R., Frank, K. A., & Krause, A. (2006). The distribution of resources and expertise and the implementation of schoolwide reform initiatives. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (Vol. 1, pp. 522–528). Mahwah, NJ: Erlbaum.

Penuel, W. R., & Gallagher, L. P. (2006). *Brief report on TIDES sample.* Menlo Park, CA: 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,* 294–315.

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). *The WHIRL co-design process: Participant experiences.* Research and Practice in Technology Enhanced Learning, *2*(1), 51–74.

Penuel, W. R., & Yarnall, L. (2005). Designing handheld software to support classroom assessment: An analysis of conditions for teacher adoption. *Journal of Technology, Learning, and Assessment, 3*(5). Retrieved September 20, 2009, from www.jtla.org

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14.

Randi, J., & Corno, L. (1997). Teachers as innovators. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *The international handbook of teachers and teaching* (Vol. 2, pp. 1163–1221). Dordrecht, The Netherlands: Kluwer.

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

Reiser, B. J., Spillane, J. P., Steinmuler, F., Sorsa, D., Carney, K., & Kyza, E. (2000). Investigating the mutual adaptation process in teachers' design of technology-infused curricula. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 342–349). Mahwah, NJ: Erlbaum.

Remillard, J. T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry, 29*(3), 315–342.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211–246.

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, NJ: Erlbaum.

Rivet, A., & Krajcik, J. S. (2004). Contextualizing instruction in project-based science: Activating students' prior knowledge and experiences to support learning. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the 6th International Conference of the Learning Sciences* (pp. 435–442). Mahwah, NJ: Erlbaum.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *Elementary School Journal, 105,* 75–102.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44,* 57–84.

Sarama, J., Clements, D. H., & Henry, J. J. (1998). Network of influences in an implementation of a mathematics curriculum innovation. *International Journal of Computers for Mathematical Learning, 3*(2), 113–148.

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

Schwartz, N. (1999). How the questions shape the answers. *American Psychologist, 54,* 93–105.

Schwartz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22*(2), 127–160.

Schwarz, C. V., Gunckel, K. L., Smith, E. L., Covitt, B. A., Bae, M., Enfield, M., et al. (2008). Helping elementary preservice teachers learn to use curriculum materials for effective science teaching. *Science Education, 92,* 345–377.

Schwarz, C. V., & Gwekwerere, Y. N. (2007). Using a guided inquiry and modeling instructional framework (EIMA) to support preservice K-8 science teaching. *Science Education, 91,* 158–186.

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

Shepard, L. (1997). *Insights gained from a classroom-based assessment project* (CSE Technical Report No. 451). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*(2), 119–150.

Shkolnik, J., Song, M., Mitchell, K. J., Uekawa, K., Knudson, J., & Murphy, R. (2007). *Changes in rigor, relevance, and student learning in redesigned high schools.* Palo Alto and Menlo Park, CA: American Institutes for Research and SRI International.

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

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

Snyder, J., Bolin, F., & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402–435). New York: Macmillan.

Songer, N. B., Lee, H.-S., & Kam, R. (2002). Technology-rich inquiry science in urban classrooms: What are the barriers to inquiry pedagogy? *Journal of Research in Science Teaching, 39,* 128–150.

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.

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

Squire, K. D., MaKinster, J. G., Barnett, M., Luehmann, A. L., & Barab, S. (2003). Designed curriculum and local culture: Acknowledging the primacy of classroom culture. *Science Education, 87,* 467–489.

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,* 963–980.

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, MA: Harvard University Press.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction.* Chicago: University of Chicago 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,* 137–158.

Wartofsky, M. (1973). *Models.* Dordrecht, The Netherlands: Reidel.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37,* 469–479.

Wehlage, G. G., Newmann, F. M., & Secada, W. G. (1996). Standards for authentic achievement and pedagogy. In F. M. Newmann & Associates (Eds.), *Authentic achievement: Restructuring schools for intellectual quality* (pp. 21–48). San Francisco: Jossey-Bass.

Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action.* Cambridge, MA: Harvard University Press.

Wertsch, J. V. (1998). *Mind as action.* New York: Oxford University Press.

Wiggins, G., & McTighe, J. (1998). *Understanding by Design.* Alexandria, VA: Association for Supervision and Curriculum Development.

Yarnall, L., Shechtman, N., & Penuel, W. R. (2006). Using handheld computers to support improved classroom assessment in science: Results from a field trial. *Journal of Science Education and Technology, 15*(2), 142–158.

Appendix
Description of Assignment Criterion 4, Scientific Inquiry

Driving Question for Coders to Consider: To what extent does the assignment engage students in all aspects of the scientific inquiry process?

Indicators: Students are asked to pose questions, select methods to use to answer those questions, carry out an investigation, and analyze and communicate results. Specifically, the assignment asks students to

- Pose an investigable question
- Decide on methods for collecting data
- Collect data
- Analyze results
- Communicate results or conclusions.

Coding

4 The assignment calls for students to plan, carry out, and communicate results of an investigation answering a student-generated question. As part of planning the investigation, students decide on methods/procedures for collecting data, collect data, analyze results, and prepare an oral or written presentation of results or conclusions.

3 The assignment calls for students to plan, carry out, and communicate results of an investigation answering a teacher- or curriculum-generated

question. As part of planning the investigation, students decide on methods/procedures for collecting data, collect data, analyze results, and prepare an oral or written presentation of results or conclusions.

2 The assignment calls for students to carry out and communicate results of a teacher- or curriculum-generated investigation. Students collect data, analyze results, and prepare an oral or written presentation of results or conclusions.

1 The assignment does not call for students to carry out and communicate results of an investigation. Students may collect data or analyze results of another investigation, but these are not part of a student investigation.

0 There is not enough information in the assignment face sheet or attached artifacts to code the assignment on this rubric with respect to the requirements for scientific communication. (Please note that you should check with the coding leader before assigning this code.)

TABLE A-1
Summary of Coding Rules for *Scientific Inquiry*

Code	*Indicators*
4	***Assignment for which all three of the following are true:*** • ***Students*** select a question to investigate • ***Students*** decide on which methods/procedures to use to investigate the question • ***Students*** collect data, analyze results, and draw conclusions
3	***Assignment for which two of the following are true:*** • ***Students*** select a question to investigate • ***Students*** decide on which methods/procedures to use to investigate the question • ***Students*** collect data, analyze results, and draw conclusions
2	***Assignment for which one of the following is true:*** • ***Students*** select a question to investigate • ***Students*** decide on which methods to use to investigate the question ***Students*** collect data, analyze results, and draw conclusions
1	**Assignment does not call for students to plan and carry out an investigation**
0	**There is not enough information to code the assignment (please see coding leader)**