American Educational Research Journal
August 2011, Vol. 48, No. 4, pp. 996–1025
DOI: 10.3102/0002831211410864

Preparing Teachers to Design Sequences of Instruction in Earth Systems Science: A Comparison of Three Professional Development Programs

William R. Penuel
Lawrence P. Gallagher
Savitha Moorthy
SRI International

This research study examined whether and how professional development can help teachers design sequences of instruction that lead to improved science learning. The efficacy of three professional development programs and a control condition was compared in a cluster randomized trial involving 53 middle school science teachers from a single school district. The four conditions varied along two dimensions: (a) the extent to which the programs guided teachers' selection of curriculum materials and (b) whether or not teachers received explicit instruction in models of teaching associated with particular methods for designing instruction. Results indicated that the two programs most effective at improving students' science learning were the ones in which teachers received explicit instruction in models of teaching.

KEYWORDS: social context, computers and learning, educational policy, networking, evaluation, instructional practices

---

WILLIAM R. PENUEL is director of Evaluation Research at the Center for Technology in Learning at SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025; e-mail: *william.penuel@sri.com*. His research focuses on teacher learning in STEM education, teacher professional communities, and the implementation of curricular reforms.

LAWRENCE P. GALLAGHER is senior research social scientist at the Center for Technology in Learning at SRI International; e-mail: *lawrence.gallagher@sri.com*. His research interests include assessment design and psychometric modeling.

SAVITHA MOORTHY is research social scientist at the Center for Technology in Learning at SRI International; e-mail: *savitha.moorthy@sri.com*. Her research interests include science education, language and literacy development, discourse analysis, media and learning, and the education of culturally diverse students, including English Language Learners.

Developing students' proficiency in science depends on teachers' skill in sequencing instructional experiences that build understanding over time (National Research Council, 2007). However, improving teachers' ability to design such sequences is not simple. It will require policies and funding that support professional development focused on developing teachers' knowledge and skill and on preparing teachers to implement models of instruction that provide students with opportunities to develop the core strands of science proficiency (National Research Council, 2007, p. 7). Evidence from empirical studies of the impacts of particular features of professional development on student learning (e.g., Fishman, Marx, Best, & Tal, 2003) indicates that policies and programs must first change teachers and teaching through content-focused professional development if student learning in science is to improve.

Professional development that prepares teachers to sequence instructional experiences for students has not been widely studied. Research is just now emerging on programs that prepare teachers to play a strong role in designing sequences of instruction for students using materials available to them or using lessons they develop. One explanation for this may be that so much emphasis in recent years has been placed on preparing teachers to follow, rather than create or adapt, curriculum materials and programs (Institute of Education Sciences, 2009). In addition, in science education, policymakers and researchers often suspect that teachers do not possess sufficient subject matter knowledge and pedagogical skill to design instruction for students, so they, too, emphasize the need to prepare teachers to follow curriculum developed by subject matter experts (Atkin & Black, 2003). Yet teachers inevitably do adapt curricula and programs to fit their classroom contexts (Squire, MaKinster, Barnett, Luehmann, & Barab, 2003), and some scholars have suggested that professional development should aim to guide teachers' design of instruction and uses of curriculum materials (M. W. Brown & Edelson, 2003; Davis & Varma, 2008).

This study investigated whether and how professional development can help teachers design sequences of instruction that lead to improved science learning. The research questions were the following:

1. Do students learn more Earth systems science when professional development guides them to select curriculum materials that are focused on learning goals when designing units of instruction?
2. Do students learn more Earth systems science when professional development for their teachers provides them with explicit instruction in models of teaching?
3. To what extent does variation in teachers' enactment of models of teaching, whether these models are taught explicitly or not to teachers, account for differences in student learning?

We investigated these questions by comparing the efficacy of three different programs and a control condition in improving Earth systems science learning in a cluster randomized trial conducted in a single school district. In the district where the study took place, district staff expected all teachers to organize Earth science instruction into 9-week units. Two of the professional development programs we studied guided teachers in their selection of curriculum materials, providing them with expert-developed units to select or adapt that were aligned to district learning goals. One of these provided explicit instruction also in the model of teaching that underlay the curriculum design to guide teachers' adaptation of their units. In a third program, teachers received explicit instruction in that same model of teaching, but they did not receive the curriculum materials the other two groups received. The control group did not receive any guidance about how to design their units of instruction or expert-developed materials, but like other teachers in the district not in the study, control teachers were free to pursue professional development opportunities on their own to help them with their unit design.

In the study, the primary outcome of interest was a measure of student learning in Earth science. The programs varied with respect to the independent variables hypothesized to affect learning, namely whether teachers were guided to select expert-developed curriculum materials and whether teachers received explicit instruction in the model of teaching linked to the expert-designed curriculum materials.

Theoretical and Empirical Context

To advance knowledge of how best to improve student learning through teacher professional development, researchers need to formulate and test hypotheses about both how particular instructional practices can improve student learning and how professional development can prepare teachers to enact those practices (Wayne, Yoon, Zhu, Cronen, & Garet, 2008). In our study, we hypothesized that for students to learn Earth system science, teachers must make effective use of instructional materials that are focused on learning goals they are expected to teach. Furthermore, we hypothesized that for teachers to use instructional materials well in the classroom, they must receive explicit instruction in the models of teaching that underlay those materials. Below, we present the theoretical frameworks and empirical research that informed the specification of these hypotheses.

Effective Use of Curriculum Materials as a Mediator of Learning

Major reform efforts in science education in the past two decades have been focused on the development of curriculum materials aligned to standards (National Research Council, 2006). Policymakers, professional organizations, and researchers view curriculum materials aligned to the goals reflected in local, state, and national standards as critical to the success of standards because developing scientific understandings reflected in standards requires students to do so in the context of engaging in the practices of science (National Research Council, 2007, 2010). Many trade textbooks and even some inquiry-oriented curriculum materials are aimed at teaching standards, but reviews suggest that few provide sufficient opportunities for students to investigate phenomena directly in a way that gives students an experience of doing science (Kesidou & Roseman, 2002). In this respect, many materials fall short of reformers' goals.

Some curricula do focus on the learning goals of standards and curriculum materials and also use innovative, research-based approaches to teaching that engage learners in scientific practices (Krajcik, McNeill, & Reiser, 2008). Studies of the efficacy of such materials provide evidence that these curricula can improve teaching and learning outcomes in science classrooms, moreover. For example, analyses of the artifacts students produced as part of learning-goals-focused, project-based science units indicate that students were engaging in complex scientific reasoning, including weighing and synthesizing many results from analyses of data and constructing scientific arguments that require synthesizing results from multiple complex analyses of data (Reiser et al., 2001). Other analyses of classroom interaction where teachers were enacting these units found that students had significant opportunities to engage in the core scientific practices of arguing from evidence and critiquing ideas and conjecturing (Tabak & Reiser, 1997). In summative evaluations in which teachers enacted instructional materials that reflect a dual focus on learning goals and innovative pedagogies and received professional development, researchers reported student performance on posttests improved significantly from pretest performance (Geier et al., 2008; Gray et al., 2001; Krajcik et al., 2000) relative to comparison students in classrooms using trade textbooks (Geier et al., 2008; Gray et al., 2001; Krajcik et al., 2000). Estimated effect sizes were moderate ($+0.37 \leq d \leq +0.44$) for larger studies that relied on standardized test scores to large ($+1.0 \leq d \leq +1.5$) for smaller scale studies that relied on researcher-constructed performance tasks closely aligned to the learning goals of the materials.

Teachers' selection and enactment of materials can and do vary in ways that can limit their efficacy. For example, teachers' methods for questioning students can limit opportunities for students to construct, elaborate on, and revise scientific explanations of phenomena they are investigating, even when curriculum materials include support for such activities (McNeill & Pimentel, 2010). Teachers' enactments may also diverge from developers' intentions for materials in the ways that teachers present science ideas, respond to student questions, structure tasks, and adapt the materials to fit their classroom contexts (Enyedy & Goldberg, 2004; McNeill, 2009; Reiser et al., 2000; Schneider, Krajcik, & Blumenfeld, 2005; Songer, Lee, & Kam, 2002). In certain circumstances, such as when teachers perceive the learning goals of curriculum materials to be incongruent with district or state standards, teachers may implement materials to a limited degree, if at all (Penuel & Gallagher, 2009; Rivet, 2006).

These findings are consistent with emerging perspectives on the importance of anticipating teachers' *uses* of curriculum in planning professional development. Learning scientists who have developed materials for science and mathematics classrooms that reflect standards and incorporate innovative pedagogies posit that designers must acknowledge and plan for the fact that teachers will and must adapt curriculum to fit their local classroom contexts (Davis & Varma, 2008; Squire et al., 2003). Remillard (2005), in her review of studies focused on teachers' uses of reform-oriented mathematics curricula, has encouraged researchers to move away from models of curriculum use that emphasize only the importance of fidelity to designers' intentions and to develop conceptual frameworks consistent with the idea that curriculum resources provide *models* (Wartofsky, 1979) for instruction and serve as *mediational means* (Wertsch, 1998) for improving student learning. 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 (M. W. Brown, 2002). From this perspective, the important questions to ask are whether teachers' intentions and adaptations to materials are productive, meaning that they do not undercut the potential efficacy of well-designed, coherent sequences of activity (Davis & Varma, 2008).

Research on teachers' curriculum use in science needs to expand to focus on how enactment of instructional practices specified by curricula mediates learning and to investigate how professional development can support teachers' enactment of those practices. More empirical research on curriculum use focuses on materials designed for mathematics classrooms, not science classrooms. Although the reform-oriented curriculum materials in both disciplines are similar in the high demands they make on teachers to transform their instructional practices (Spillane, 1999), more research needs to focus on how teachers' enactment of instructional practices demanded by curricula mediate student learning in science classrooms. Furthermore, researchers in science education who study science curriculum use are now only beginning to test their recommendations for how to organize professional development for productive adaptations. Below, we review those recommendations and findings from studies of programs that adhere to those recommendations and locate our own study within this emerging area of research on teacher professional development.

The Need for Explicit Instruction in Models of Teaching Specified by Curriculum

One recommendation that has emerged from studies of teachers' adaptations of science curriculum materials is that professional development programs should provide teachers with explicit guidance or instruction in the models of teaching specified within materials. For inquiry-oriented materials in particular, teachers need to know how to be systematic in using inquiry-oriented methods to serve a focused instructional purpose, including providing guidance and structure to students' encounters with phenomena. Some researchers have recommended that supports be embedded within the materials themselves, arguing that curriculum needs to provide teachers with detailed lesson descriptions and ways of responding to student thinking (Davis & Krajcik, 2005; Schneider et al., 2005; Schneider & Krajcik, 2002). Preparing teachers to elicit and address the ways students reason about phenomena is especially important in science because problematic ideas and misconceptions can inhibit growth in students' understanding of concepts (Minstrell & Kraus, 2005). Other researchers call for science educators and professional development leaders to make explicit the intentions of particular activities for teachers and rationale behind their sequence (Lin & Fishman, 2006). These researchers point to evidence that teachers often have difficulty understanding the principles underlying designs and how designers of materials intend specific lessons to connect with one another to create a coherent curriculum sequence (Lin & Fishman, 2006).

Embedding supports into materials is unlikely to be sufficient for addressing ways that teachers' instruction could limit the efficacy of curriculum materials, a point that the researchers who have made these recommendations readily concede (see, e.g., Davis & Krajcik, 2005). Teachers do not always refer to or use those aspects of materials intended to support their own learning, and when they do, teachers (as all readers do) make sense of the content of the materials in light of their own interpretive frames for teaching (McNeill, 2009; Remillard, 2000; Schneider & Krajcik, 2002). Furthermore, teachers may find it difficult to discern strategies for engaging students in scientific discourse from the written materials alone (Alozie, Moje, & Krajcik, 2010). Professional development that provides opportunities for active learning, that is, direct experience with the kind of learning environment they are expected to construct, may be needed for teachers to adhere to the models of teaching specified in curriculum (Marx & Harris, 2006).

Though it is increasingly a standard practice within science teacher professional development to immerse teachers in the inquiry-oriented curricula they are about to enact (see, e.g., Borman, Gamoran, & Bowdon, 2008), the potential importance of providing *explicit instruction* in the models of inquiry teaching alongside this immersive experience is only now beginning to be recognized within the field. The need for both immersion and explicit or overt instruction for deep understanding is a principle for the design of learning environments that is supported by a large body of research on how people learn (National Research Council, 1999). On the one hand, immersive experiences can provide learners with a rich context that motivates their learning (Barron et al., 1998; Blumenfeld, Soloway, Marx, Guzdial, & Palincsar, 1991). On the other hand, when learners encounter models or phenomena only through immersion, especially when they encounter only a few examples, learners may not attend to salient features of models, or they may overgeneralize what they take to be the lessons of their immersive experiences (Barron et al., 1998; Bransford & Schwartz, 1999).

A potential way forward for professional development programs is to provide explicit instruction to teachers aimed at preparing them to plan to use particular pedagogical strategies when designing instruction, either with or without supporting curriculum materials. One such program of professional development is *Understanding by Design* (UbD; Wiggins & McTighe, 1998). UbD is a framework for designing curricular units of instruction that centers on the big ideas, essential questions, and authentic performances (Wiggins & McTighe, 1998). The approach, as articulated by its developers, calls for designers to follow three steps in sequence: (a) articulate learning goals; (b) design, select, or adapt assessments to measure student progress toward those goals; and (c) design, select, or adapt activities that will develop student understanding as measured by the assessments. According to the UbD model, the learning goals should focus on developing enduring understandings of concepts that are core to a discipline. Teachers following the approach should make judgments of students' understanding on the basis of assessments of their performance in authentic tasks that require applying knowledge and skills developed over the course of a unit (McTighe & Wiggins, 1999). To ensure that students perform well on such assessments, teachers need to use a variety of teaching strategies that elicit and develop different facets of student understanding, which include the ability to explain and apply a concept to answer a question or solve a difficult problem (J. L. Brown, 2004). Underlying UbD's emphasis on organizing instruction so that students gain practice in developing explanations and applying knowledge is the theory of learning that students must have opportunities to construct knowledge to develop depth of understanding, not just read it from a book or hear it from a teacher (McTighe & Seif, 2003).

The UbD models of pedagogical practices that are intended to support teachers' design share some features with other widely used models for unit design in science education. These include both project-based learning (Blumenfeld et al., 1991; Krajcik & Blumenfeld, 2006; Krajcik & Czerniak, 2007; Krajcik et al., 2008) and the 5E (Engage–Explore–Explain–Elaborate–Evaluate) instructional model (Bybee, 1997, 2004; Bybee et al., 2006). For example, strategies and exemplars for identifying "where to and where from" (Wiggins & McTighe, 1998, p. 200) support the elicitation of student preconceptions in UbD and serve a similar purpose to strategies for engagement in the 5E model. The strategy of "hooking and holding" (Wiggins & McTighe, 1998, p. 201) in UbD serves a similar purpose as the introduction of a motivating problem in project-based learning, namely, to immerse students in a problem that makes learning goals immediate, compelling, and concrete. Although UbD differs from these and other models in the specific guidance for teachers in unit design (Krajcik et al., 2008), selected programs (including two we studied) provide more detailed guidance on pedagogical strategies linked to specific subject matter domains. For example, the *Earth Science by Design* program (McWilliams et al., 2006) shows teachers how to incorporate visualizations into instruction because interpreting and using visualizations are important practices in Earth systems science.

The Current Study

This article reports results from an efficacy study in a single large urban district in the southeastern United States. The conditions in the participating district were ideal for our study because (a) the district had recently adopted a UbD approach for its teachers to use in guiding their development of four 9-week units during the school year and (b) the district had not yet invested significant professional development in preparing teachers to use the approach. This enabled us to create experimental conditions in which teachers were supported in designing units in contrasting ways.

Researchers, in consultation with the district, selected three professional development programs to compare in the study: *Investigating Earth Sciences* (*IES*; Condition 1), *Earth Sciences by Design* (*ESBD*; Condition 2), and the *Hybrid* (Condition 3). Teachers who volunteered but were assigned to design units according to the district's requirements without professional development or additional materials constituted the control (Condition 4). There was a requirement of all teachers in the district that they organize their Earth science instruction into a 9-week unit. The four conditions differed along two dimensions: (a) whether teachers received professional development in which they were guided to select materials focused on learning goals and that incorporated inquiry-oriented pedagogy and (b) whether the teachers received professional development that provided them with explicit instruction in models of teaching. Figure 1 depicts how each condition represents the experimental contrasts of interest in the study.

Earth Sciences by Design (Condition 1)

ESBD is a yearlong program of professional development created by TERC and the American Geological Institute (AGI) with funding from the National Science Foundation (NSF).[1] *ESBD* prepares teachers to apply the principles of UbD to teach Earth system science. In this condition, teachers were the designers of instructional experiences for students and received explicit instruction in the models of teaching and assessment associated with UbD. The teachers were instructed in the UbD approach and were required to apply it to plan a 9-week Earth science unit and using curricular materials for students they selected. No specific curriculum was recommended to the teachers; instead, they were expected to reorganize existing curricular materials, such as those from their textbooks or those they may have developed themselves or collected from colleagues at professional

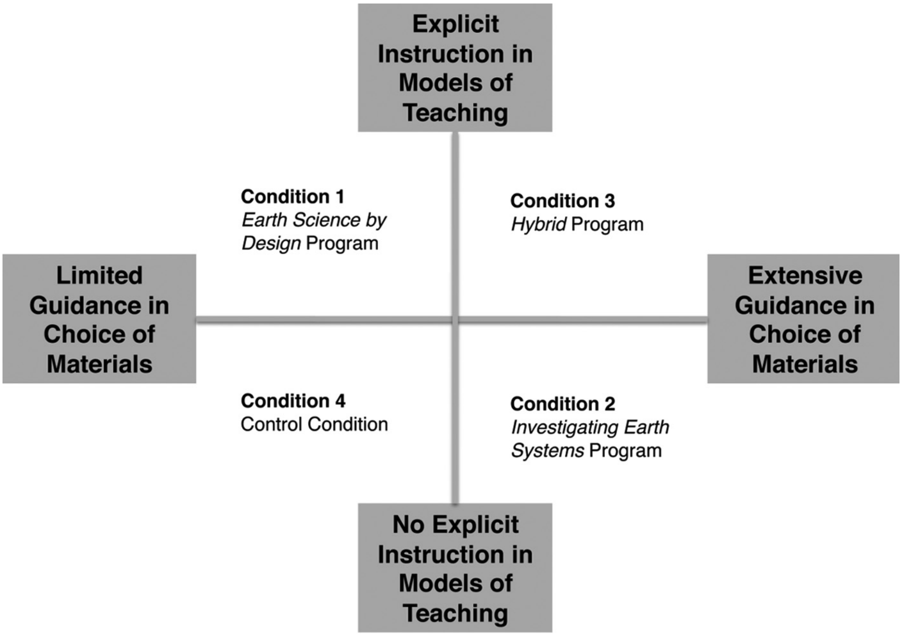

Figure 1. **Experimental contrasts and the four conditions.**

conferences, into coherent units of instruction that targeted essential questions and enduring understandings and culminated with a performance assessment. As part of their professional development, teachers assigned to the *ESBD* condition received explicit instruction in the pedagogical strategies of UbD, as adapted by the TERC team for Earth systems science teaching and learning. Hands-on practice in the *ESBD* condition's professional development took the form of trying out visualizations and engaging in unit planning.

Investigating Earth Sciences (Condition 2)

IES is a 10-module middle school curriculum, funded by the NSF and developed by AGI. It has the specific purpose of preparing teachers to implement specific modules in a manner that fits a school district's middle school Earth science standards. The training for teachers in this condition was intended to immerse teachers in the types of activities of the *IES* curriculum, especially the student investigations), with the assumption that the teachers would implement instructional sequences using the materials as laid out. Furthermore, the teachers were expected to limit their use of materials to the *IES* curriculum as they developed and delivered their units, and they were explicitly asked to refrain from using other materials except the resources that were available on the *IES* website. Thus, their design of units was highly constrained and consisted primarily of establishing a pacing guide for their units. Importantly for purposes of the study, teachers in the *IES* condition did not, as part of their professional development, receive explicit instruction in the models of teaching that underlay *IES*, even though this curriculum did incorporate pedagogical and assessment strategies that reflected the UbD approach.

Hybrid (Condition 3)

The *Hybrid* condition was a blend of the *IES* and *ESBD* conditions. Teachers assigned to this condition received explicit instruction in the models of teaching that are at the heart of the UbD approach, and they also were immersed in the practice of design as in the *ESBD* condition (Condition 1). However, rather than developing instructional experiences by flexibly assembling materials from a wide variety of sources, teachers were required to draw at least 50% of their curricular materials from the *IES* modules relevant to their grade level rather than compiling all their materials from a wide variety of sources. Hands-on practice in the *Hybrid* condition activities took the form of practicing *IES* investigations and engaging in unit planning. Throughout all professional development activities, UbD concepts underlying the design of the *IES* materials were emphasized. This particular condition provided the research with an experimental condition in which teachers received both extensive guidance about what materials to select and explicit instruction in the pedagogical strategies of UbD. In addition to the pedagogical and assessment strategies explicitly taught in *ESBD*, in the *Hybrid* program teachers gained practice with analyzing expert-developed curriculum materials for evidence of application of those strategies that they were expected to use in class.

Control (Condition 4)

The *Control* condition featured a strong element of design, but the teachers did not receive explicit instruction in the models of teaching underlying UbD, nor did they receive materials that reflected those models. Instead, the district standards, which had been written in accordance with the UbD approach, guided teachers' planning of instructional sequences. Although teachers had the option to participate (or not) in any district-offered professional development, they did not participate in the professional development programs associated with the study.

Commonalties and Differences Among the Programs

It was important to the integrity of our study that the programs shared features essential for high-quality professional development and that they differed only in their support for teachers learning to *design sequences of instruction using models* in Earth science. Table 1 summarizes the commonalties among the programs relative to the five core features of professional development identified by Desimone (2009).

The three programs did differ in how they prepared teachers to design sequences of instruction for students, the degree to which they offered explicit instruction in models of teaching, and the exemplars provided to support teaching following the UbD model. Table 2 shows how the three programs and the *Control* condition differed. The chief differences pertained to whether teachers were prepared, through explicit instruction in models of teaching, to design curricular units (*ESBD*, *Hybrid*) or not (*IES*, control).

Relating the Research Questions to Conditions and Hypotheses

Comparing learning gains of students whose teachers were assigned the *IES* and *Hybrid* conditions (Conditions 2 and 3) to gains made by students in *ESBD* and *Control* classrooms addresses the question of the role of guiding teachers in the selection of curriculum materials for use in their Earth systems science units (Question 1). The *IES* and *Hybrid* teachers' design activities were constrained, in terms of the choices they had for making decisions about what curriculum materials to use; if their students outgained those of their colleagues who were assigned to the *ESBD* and *Control* condition, those data would provide evidence to support the importance of guiding teachers' selection of curriculum materials.

Comparing learning gains of students whose teachers were assigned the *ESBD* and *Hybrid* conditions (Conditions 1 and 3) to gains made by students in *IES* and *Control* classrooms addresses the question of the role of providing teachers with explicit instruction in the instructional models associated with the UbD approach to unit design (Question 2). If students from *ESBD* and *Hybrid* classrooms outperformed students in *IES* and *Control* conditions, results would provide evidence to support the importance of providing explicit instruction in models of instruction as part of professional development.

Because the *Hybrid* professional development program provided both guidance with respect to curriculum and explicit models of instruction, we expected students in classrooms whose teachers were assigned to this condition to outperform all other groups in the study (Hypothesis$_1$; H_1). We hypothesized that students the *ESBD* and *IES* conditions would outperform the control group classrooms, but score somewhat lower than students in the *Hybrid* classrooms, since their teachers had been given one but not both kinds of support that research indicates may be needed to promote effective curriculum use (Hypothesis$_2$; H_2). An alternate hypothesis that informed the overall design and sample size required for the study was that there would be no differences among any of the groups with respect to student learning gains (Hypothesis$_{1\text{-}2\text{-}0}$; $H_{1\text{-}2\text{-}0}$).

Table 1
Features the Three Professional Development Programs Have in Common

Feature	*ESBD*	*IES*	*Hybrid*
Support for subject matter learning and learning about student conceptions	Exploration of "big ideas" as defined by Earth systems scientists	Discussion of "key ideas" at the beginning of each investigation	Exploration of "big ideas" as defined by Earth systems scientists
	Teachers reflect on common student conceptions of big ideas	Teachers' guide includes typical student conceptions	Teachers reflect on common student conceptions of big ideas
Use of active learning strategies	Practice with selecting enduring understandings, assessments, and materials	Practice with implementing investigations	Practice with selecting enduring understandings, assessments, and materials
	Group discussions	Group discussions	Group discussions
	Creating a 9-week unit	Creating a 9-week pacing guide	Creating a 9-week unit
Collective participation	Teachers participate in discussion and unit development as grade-level teams within the workshop	Teachers participate in discussion and practice investigations as grade-level teams within the workshop	Teachers participate in discussion and unit development as grade-level teams within the workshop
Duration and time span	2-week summer workshop, 4 days of follow-up throughout the year	2-week summer workshop, 4 days of follow-up throughout the year	2-week summer workshop, 4 days of follow-up throughout the year
Coherence with teachers', schools', and districts' goals for teacher learning	Program supports the teachers in designing units aligned with district goals	*IES* curriculum modules aligned with district standards	Program supports the teachers in designing units aligned with district goals

Question 3 addresses specifically the possibility that differences in learning gains have no relationship to instruction as enacted. In other words, there might have been evidence to support H_2 but no evidence that teachers in these conditions engaged more frequently in the target pedagogical practices of UbD or any evidence of a link between these practices and student

Table 2
Distinctive Features of the Four Study Conditions

	ESBD	*IES*	*Hybrid*	*Control*
Role of program in preparing teachers to design sequences of instruction	Teachers learn *Understanding by Design* (UbD) approach to designing units of instruction	None: Teachers develop pacing guide for sequence established by experts	Teachers learn UbD approach to designing units of instruction	Teachers design units following the UbD approach but receive no special training in the approach
Models and materials available for unit design	Models of instruction from UbD	Not applicable	Models of instruction from UbD	Textbook
	Textbook		*IES* curriculum materials	Teacher-found and teacher-made lessons
	Teacher-found and teacher-made lessons		Textbook	
			Teacher-found and teacher-made lessons	
Degree of explicitness in teaching UbD model of instruction	High: Teachers encounter strategies for developing and assessing student understanding in multiple sessions	None	Medium: Teachers encounter strategies for developing and assessing student understanding in 2–3 sessions	None
	Pedagogical strategies emphasized are WHERETO strategies		Pedagogical strategies emphasized are WHERETO strategies	
	Assessment strategies emphasized are preconception assessments, performance tasks, and student self-assessment		Assessment strategies emphasized are preconception assessments, performance tasks, and student self-assessment	

learning (Hypothesis$_{3\text{-}0}$; $H_{3\text{-}0}$). A mediation analysis (described below) focused on the specific pedagogical strategies taught in the professional development addresses the question of whether being assigned to a particular treatment group is associated with enacting those strategies. We hypothesized that enactment of these strategies would mediate learning gains of students (Hypothesis$_3$; H_3). Although a mediation model cannot be used to establish strict causality, it provides an important additional source of evidence regarding what supports for design of instructional units are particularly critical to provide as part of teacher professional development.

Procedure for Random Assignment

Teachers were randomly assigned to one of the three professional development conditions or to the *Control* condition. Our randomization scheme had three 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).

Power Analysis

In planning the efficacy study, we conducted a statistical power analysis to determine the necessary sample size for a multisite cluster randomized trial using the statistical program Optimal Design (Liu, Congdon, & Raudenbush, 2001). Earlier studies of the impact of teacher professional development on children's science learning indicated the need to plan for medium to large effects (Kennedy, 1999). We assumed an intraclass correlation coefficient of .15, 25 students per classroom, a pretest–posttest correlation of .70, and a desired power of .80. We aimed to recruit four teacher participants each from 15 different schools (n = 60), which could detect effects of .24 or larger.

Participants

A total of 53 sixth, seventh, and eighth grade teachers from 19 middle schools in a large urban district participated (see Table 3). Of these, 13 teachers were assigned to the *ESBD* condition (the program to prepare teachers to design their own units), 13 to the *IES* condition (the program to prepare teachers to adopt a high-quality Earth science unit), 13 to the *Hybrid*

Table 3
Characteristics of Faculty Respondents to Questionnaire

	Condition			
	IES	*ESBD*	*Hybrid*	*Control*
Gender, percentage				
Male	23	38	23	29
Female	77	62	77	71
Race/ethnicity, percentage[a]				
White	77	46	46	79
African American	15	46	46	14
Hispanic or Latino	8	0	15	7
Asian	8	0	15	0
Other or unknown	0	8	8	0
Teaching experience, years				
Teaching				
M	13.2	14.9	8.4	12.8
SD	11.8	11.8	8.4	9.0
Teaching science				
M	11.5	8.8	4.3	10.4
SD	9.6	6.1	3.3	8.1
Highest degree, percentage[b]				
Bachelor's	69	77	85	64
Master's	23	23	8	36
Educational specialist's	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.
[a]Teachers could select multiple categories.
[b]Totals may not sum to 100% because of rounding.

condition (the program designed to prepare teachers to design their own units using a combination of high-quality Earth science units and other materials), and 14 to the control group. There were no significant differences among groups on gender, race/ethnicity, teaching experience, highest academic degree, or grade assignment. On average, teachers in the *ESBD* condition had less experience teaching than teachers in the other conditions, and teachers in the *Hybrid* condition had less experience teaching science than teachers in the other three conditions. The differences among groups were not statistically significant for overall years of teaching and were marginally nonsignificant (χ^2 = 7.59, df = 3, $p < .06$) for years teaching science.[2]

Table 4
Characteristics of Student Participants

	Condition			
	IES	*ESBD*	*Hybrid*	*Control*
Number of student participants				
Grade 6	156	267	178	85
Grade 7	86	50	94	124
Grade 8	157	102	43	208
Total	399	419	315	417
Gender				
Percentage female	57	58	53	55
Percentage male	43	42	47	45
Race/ethnicity[a]				
Percentage White	56	51	49	50
Percentage African American	22	31	35	33
Percentage Hispanic	10	7	6	10
Percentage Asian or Pacific Islander	8	8	4	5
Percentage Native American	3	5	3	3
Percentage Other	9	10	13	7

[a]Percentages are based on 1,291 out of 1,550 student who indicated at least one of the listed races/ethnicities. Students who omitted any response to these items were not included in the calculation of percentages.

Although members of every group had received some professional development in the UbD approach, there were no significant differences among groups in overall exposure. Similarly, there were no significant differences among groups in professional development related to the development of classroom assessments or unit planning.

These teachers taught a total of 1,550 students (see Table 4 for key characteristics of students in the sample). There were no significant differences in gender, race/ethnicity, or grade level distribution of students by experimental condition.

Measures

Classroom observations. We constructed an observation protocol to measure the extent to which teachers' instruction was aimed at teaching for deep understanding in science. Observers marked the extent to which teachers engaged in each of the following strategies aligned with the UbD "where to" and "where from" strategies for organizing instruction described in the introduction to the article: Students had multiple opportunities to reflect on what they were learning; students were knowledgeable about the purpose of the activity and understood why it was being taught; a variety of instructional activities (three or more) were used to teach a single concept; teachers tailored lessons to individual students' needs, skills, and interests; teachers actively engaged students in the activities of the class; and teachers elicited and made use of student preconceptions or ideas about the topic in their lessons. For each item, observers marked whether the descriptor was *not at all true of this class*, *somewhat untrue of this class*, *somewhat true of this class*, or *very true of this class*.

A researcher visited each classroom once during the implementation of teachers' Earth science unit. Visits took place in the middle of units, within a 1-week period (all teachers were in roughly the same place in their units, ensuring comparability of observations across classrooms). All observers had been trained in the protocol and had practiced using it in the field with classrooms not in the study. In addition, while in the field observers conducted 15 observations as pairs to allow for calculation of the percentage of interrater agreement. For facets of understanding, interrater agreement was between 80% and 100%. Ratings of the two observers on their judgments of students' understanding of the purpose of the instructional activity were exactly the same in 66.7% of the observations and within one point in 100.0% of the observations. Ratings on judgments about the extent to which teachers elicited student preconceptions were the same in 86.7% of observation and within one point in all observations.

Teacher postunit survey. On completing the science unit, teachers completed a survey administered online. The survey was divided into sections probing for supports and hindrances to implementation, the average length of classes and units completed, attitudes toward study participation, and the motivation behind particular pedagogical practices. The last category included the use of visualizations, the reuse of activities from the prior year, and the use of a variety of assessment techniques. In this article, we focus on the use of three assessment techniques specified in the professional development programs: (a) preconception assessments (specified in all three programs), (b) performance assessments (specified in the *ESBD* and *Hybrid* programs), and (c) student self-assessment (specified in the *ESBD* and *Hybrid* programs).

Using a commercial web-based survey program, the research team administered the postunit questionnaires to teachers online. Teachers completed the questionnaires within 3 to 4 weeks after they finished teaching their Earth science units. We used the online survey program to monitor response rates and follow up with teachers; we obtained a 91% response rate (n = 53).

Unit assessments. A team of SRI International researchers developed the unit tests. There were three versions aligned with Earth science content standards taught in each grade: the Dynamic Earth for Grade 6, Water on

Earth for Grade 7, and the Final Frontier (Astronomy) for Grade 8. The three versions were administered both as pretests and as posttests. Items were classified into three types: recall (demanding that students recall a fact, definition, or concept that was part of the standard), explanation (requiring students to recognize a correct explanation for a scientific phenomenon), and application (requiring students to apply what they had learned about a concept to a particular context or situation). In addition to the content assessment, each test contained a brief survey of student background characteristics, including gender, age, and grade.

Construct validity evidence was developed by having an expert panel of district staff judge whether each test item was aligned with the grade-specific district standards in Earth science. Items rated by a team of three district staff members as having low alignment were excluded from pretesting. Pretest booklets were created, containing approximately 35 items, and administered to a small sample of students to assess the appropriateness of the wording, difficulty level, and test length. On the basis of an analysis of item difficulties and classical item fit statistics, final instruments of 21 items (Grade 6) and 22 items (Grades 7 and 8) were created.

After the pretesting and posttesting of the unit tests, item response theory (IRT) scaling methods were used to estimate ability scores for each student as well as for post hoc analysis of item characteristics. First, two-parameter logistic models were fit to the posttest response data at each grade level using the MULTILOG software package (Thissen, Chen, & Bock, 2003). This resulted in characterization of each item by two parameters: item discrimination (the degree to which the probability of correctly responding to an item varies with the level of student knowledge) and item difficulty. These item characteristics were subsequently used to estimate scores for students on both the pretest and posttest.[3] An analysis of item parameters showed that nearly all items exhibited desirable discrimination parameters and a moderate range of difficulty. The one or two items in each test with low discrimination parameters were retained in the score estimations to avoid selectively dropping items insensitive to these particular curriculum units. The classical Cronbach's alpha measures of internal consistency for the complete posttests ranged from .75 to .78.

Approach to Analysis

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, students were nested within teachers (who were part of professional development conditions) and teachers were nested within schools.

Direct effects model (Model 1). To address our question about the direct impact of the experimental conditions on student learning, we fit the following three-level model to the data from the student learning assessment, collapsing across the three grade levels.

Level 1 (student)

$$Y_{ijk} = \pi_{0jk} + e_{ijk} \tag{1}$$

Level 2 (teacher)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}IES_{jk} + \beta_{02k}ESBD_{jk} + \beta_{03k}Hybrid_{jk} + r_{0jk} \tag{2}$$

Level 3 (school)

$$\begin{aligned} \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{01k} &= \gamma_{010} \\ \beta_{02k} &= \gamma_{020} \\ \beta_{03k} &= \gamma_{030} \end{aligned} \tag{3}$$

In this model *ESBD*, *IES*, and *Hybrid* are 0 or 1 indicator variables for the respective experimental conditions (the *Control* condition being the omitted category). The outcome Y_{ijk} is a standardized student gain score (computed by subtracting the pretest from the posttest score and dividing by the standard deviation of the grade-specific pretest). The variables u_{OOk}, r_{Ojk}, and e_{ijk} are random error terms at the school, teacher, and student levels, respectively.

Mediation model. To address our second question about the specific mediating effects of particular teacher-level variables, we fit the following trio of models to the student learning data. We focused on two constructs central to the UbD model in our analysis: teachers' use of the WHERETO strategies in teaching and their use of assessment strategies in the model. We did not incorporate the third main feature of UbD, the specification of enduring understandings (goals for student learning), because teachers did not have to establish these. We did not, therefore, measure teachers' articulation of enduring understandings or measure their use.

Following classic mediation analysis (Baron & Kenny, 1986), we modeled (a) the direct effects of the experimental condition on the mediator, (b) the direct effect of the mediator on student learning, and, should both these models show significant effects, (c) the joint effects of experimental condition and mediator on student learning. Because all mediators are at the teacher level (i.e., a 2-2-1 model; see Bauer, Preacher, & Gil, 2006), the first model excluded student-level variables. The three models are presented below in collapsed form.

Model 2 (effects of condition on teacher mediator)

$$M_{jk} = \gamma_{000} + \gamma_{010}IES_{jk} + \gamma_{020}ESBD_{jk} + \gamma_{030}Hybrid_{jk} + u_{00k} + r_{0jk} \tag{4}$$

Table 5
Descriptive Statistics for Student Pretest, Posttest, and Gain Scores, by Condition

	Experimental Condition				
	ESBD	*IES*	*Hybrid*	*Control*	Overall
Pretest					
M	469	480	453	463	467
SD	80	84	83	83	83
Posttest					
M	547	532	527	514	530
SD	100	94	111	105	103
Gain					
M	78**	52	74*	50	63
SD	77	74	86	80	80

*$p < .05$ (comparison to *Control* condition). **$p < .01$ (comparison to *Control* condition).

Model 3 (effects of mediator on student learning)

$$Y_{ijk} = \gamma'_{000} + \gamma'_{010}M_{jk} + u'_{00k} + r'_{0jk} + e'_{ijk} \tag{5}$$

Model 4 (joint effects of experimental condition and mediator on student learning)

$$Y_{ijk} = \gamma''_{000} + \gamma''_{010}IES_{jk} + \gamma''_{020}ESBD_{jk} + \gamma''_{030}Hybrid_{jk} + \gamma''_{040}M_{jk} + u''_{00k} + r''_{0jk} + e''_{ijk} \tag{6}$$

Results

Impacts on Student Learning

Table 5 presents descriptive statistics for mean pretest, posttest, and gain scores by condition. The omnibus hypothesis of group equivalence of pretest scores was tested and found to be nonsignificant.

Expressed as a standardized effect size (using the pooled pretest standard deviation of 83), the *ESBD* and *Hybrid* conditions showed greater gains than the control group, with standardized effect sizes of .34 and .29, respectively. A version of Model 1 (using raw achievement scores) was fit, and these gains were found to be statistically significant ($z = 2.73$, $p < .01$, and $z = 2.54$, $p < .05$, respectively). This evidence supports the idea that professional development providing explicit instruction in models of teaching can increase student learning in Earth system science. Just as important, the learning gains under the *IES* condition did not significantly differ from those under the *Control* condition; thus, the study findings provide mixed evidence for the claim that guiding teachers to choose standards-aligned curricula that also embed opportunities for student investigations of phenomena is

Table 6
Means of Rating Scales for *Understanding by Design* Pedagogical Strategies

	IES	*ESBD*	*Hybrid*	*Control*
Students had multiple opportunities to reflect on what they were learning.	2.3	2.3	1.8	1.4
Students were knowledgeable about the purpose of the activity and understood why it was being taught.	2.6*	2.6*	2.4	2.0
A variety of instructional activities (3 or more) were used to teach a single concept.	1.3	2.2*	1.4	1.5
Teachers tailored lessons to individual students' needs, skills, and interests.	2.0	1.8	1.3	1.6
Students were engaged in the activities of the class.	0.7	1.1	0.2	0.3
Teachers elicited and made use of student preconceptions or ideas about the topic in their lessons.	0.9	0.9	1.1	0.6
Overall composite	1.6	1.8*	1.4	1.2

Note. Scale: 0 = *not at all true of this class*, 1 = *somewhat untrue of this class*, 2 = *somewhat true of this class*, 3 = *very true of this class*.
*Statistically significantly different from *Control* condition after controlling for false discovery rate of .05 across 21 comparisons.

likely to produce greater learning gains. Although students in the *Hybrid* condition did perform well relative to both the *IES* and *Control* conditions, they did not outperform students in the *ESBD* group. Thus, we found mixed evidence for H_1 and H_2; however, the results did not support $H_{1\text{-}2\text{-}0}$, the hypothesis that there would be no differences among the groups.

Classroom Practices

We examined the potential effects of two different classroom practices that are design principles for instruction within UbD and that were explicitly taught in the two professional development programs found to be effective: teachers' use of the pedagogical strategies for developing students' understanding and teachers' use of a range of assessment strategies intended to gather evidence of students' growth in understanding of science concepts being taught. We chose these two sets of practices because they are the heart of UbD's models for what teaching in a "UbD classroom" should look like. The pedagogical strategies correspond to the "where to" and "where from" strategies in UbD (Wiggins & McTighe, 1998, chap. 9), and the assessment strategies to guidance provided in the model for "Thinking Like an Assessor" (Wiggins & McTighe, 1998, chap. 7).

Pedagogical strategies. Trained observers rated teachers' use of strategies corresponding to the pedagogical principles of UbD using the criteria listed in Table 6. Each row of the table was coded on a scale from 0 (*not at all true*

Table 7
Coefficients of Mediation Model With Pedagogical Strategies as a Mediator

Predictor	Student Learning by Condition (Model 1)	Mediator by Condition (Model 2)	Student Learning by Mediator (Model 3)	Student Learning by Condition, Controlling for Mediator (Model 4)
IES	0.047	0.796*	—	−0.113
ESBD	0.459**	1.083**	—	0.243
Hybrid	0.465**	0.182	—	0.426**
Pedagogical strategies	—	—	0.172*	0.193**

Note. Pedagogical strategies scaled to unit variance across teachers.
*$p < .05$. **$p < .01$.

of this class) to 3 (*very true of this class*), and a standardized composite score was computed from the mean of these six ratings ($\alpha = .66$).

Table 7 below shows results for each of the models fit to the data. Model 1 shows the main effects analysis, with each condition's effect compared to the control group. Model 2 shows the effect of the mediator on student learning by condition, and Model 4 shows the impact on student learning by condition, controlling for the mediator. The mean rating of 1.8 for the *ESBD* classrooms is significantly greater than that of the control classrooms when Model 2 above is fit ($z = 2.99$, $p < .01$). When Model 3 is fit to ascertain the power of the pedagogical strategies composite to predict learning, the standardized coefficient is .172 ($z = 2.69$, $p < .01$). When combining the experimental conditions and pedagogical strategies (Model 4), we see that the magnitude of the *ESBD* effect on student learning decreases to nearly half its original value (becoming nonsignificant; see Table 7). This suggests that pedagogical strategies may serve as a mediator for the *ESBD* condition, providing evidence supporting H_3.

Assessment techniques. The postunit survey asked about assessment techniques used during the units. Teachers responded with a simple yes or no, indicating whether they used a particular technique during the Earth science units. The three techniques specified in at least one of the professional development models and the proportion of yes responses are listed in Table 8.

One practice explicitly specified within the *ESBD* and *Hybrid* programs is the use of a preconception test. This serves as an early formative assessment of student understandings and aids the teacher in directing instruction toward particular misconceptions or gaps in student knowledge. Of *ESBD* teachers, 92% used this assessment technique, compared to 29% of the control teachers, a statistically significant difference ($p < .001$). In addition, 77%

Table 8
Percentage of Teachers Using Specific Assessment Techniques, by Condition

	Experimental Condition (%)			
Assessment Technique	*IES*	*ESBD*	*Hybrid*	*Control*
Preconception test or quiz	8	92*	77*	29
Student self-assessment	15	38	46	21
Performance tasks	92	92	85	79

*Statistically significant after controlling for false discovery rate of 5% across the 9 possible comparisons with the *Control* condition.

of the *Hybrid* teachers used this technique as well, a statistically significant difference from the control group teachers ($p < .01$). There were no differences between the control group teachers and teachers in any of the programs with respect to use of self-assessments or performance tasks.

The use of preconception tests (embedded assessments, not to be confused with pretests used in the impact analysis) also significantly predicts standardized learning gains ($\beta = .29$, $z = 2.47$, $p < .05$). However, when entered as a mediator in Model 4, the coefficient becomes nonsignificant, and the coefficients for the *ESBD* and *Hybrid* conditions are largely unchanged. This suggests that the use of preconception testing is an indicator of a teacher being in the *ESBD* condition (it is predicted by *ESBD* and as an indicator predicts student learning when used as a sole predictor). We do not yet have evidence that preconception testing is causally related to learning gains; thus, evidence from analysis of teachers' reported assessment strategies does not support H_3.

Discussion and Conclusions

Our study findings provide strong evidence for the potential efficacy of professional development programs that prepare teachers to design sequences of instructional experiences for students, provided they present teachers explicit models of teaching and assessing learning to support their design. Students from teachers who participated in both the *ESBD* and *Hybrid* programs outperformed students of teachers assigned to the *Control* condition. Both of these programs cast teachers in a role that gives them responsibility for designing instructional experiences using a combination of expert- and teacher-designed activities.

The study provided more mixed evidence regarding the importance of providing strong curricular guidance for helping teachers design units of instruction in Earth systems science. On the one hand, students in the *Hybrid* condition outgained students in the *Control* condition, a group who did not benefit from encounters with the *IES* materials. However, the students in *Hybrid* classrooms made gains that were not statistically different from those in the *ESBD* classrooms, and the *IES* students made gains that were not statistically different from those of *Control* students.

The mediation analysis partially helps to explain these findings: This analysis points to the importance of pedagogical strategies taught as part of the *ESBD* program for improving student learning. For the *ESBD* condition, teachers' enactment of these strategies mediated students' science learning. Significantly, teachers in this condition had the most explicit instruction in these strategies. Though nonsignificant, the direction of mediation effects was in the expected direction for the *Hybrid* students. On the basis of the mediation analysis, we conclude that a key feature of the professional development that may have made it effective was providing teachers with explicit guidance in applying models for instruction following the principles of UbD. A corollary claim not directly tested in the study, but implied by the mediation analysis, is that engaging teachers as designers of instructional sequences without such guidance in models of instruction might not be effective in promoting student learning.

It would be wrong to conclude from the study, however, that guiding teachers to use curriculum materials that are aligned to standards and that provide students with direct encounters with scientific phenomena through investigations is not a useful professional development strategy. Though teachers and the district initially reported that from their perspective, the professional development and curriculum materials were just as well aligned to district standards as those in the other study conditions (see Penuel & Gallagher, 2009), our own reanalysis of the alignment of standards to the curricular units chosen for the study suggested to us that a few concepts on the district standards got less adequate treatment in *IES* than might have been available, for example, through even the district-adopted textbook. A clear challenge for curriculum developers, then, is the degree to which the patchwork of local standards inhibits creating materials that are strongly aligned to learning goals of different localities, including the one in this study. Though teachers indicated on postunit surveys that they did in fact vary in their coverage of topics included in the unit tests, topic coverage did not vary systematically by condition, and there were no significant associations between weeks topics were covered and student gains.[4]

The study findings do add to what we know about effective professional development. Past research on professional development has focused on a number of features common to effective professional development, but not on teachers' role in curriculum design (Desimone, 2009). Furthermore, the study supports the argument made by some (e.g., Alozie et al., 2010) that providing materials alone to teachers, no matter how supportive of teacher learning in and of themselves, is not sufficient to ensure changes to teaching and learning. Rather, consistent with what a number of scholars have argued (e.g., Brown, 2009), what is particularly important is that teachers develop the capacity to design sequences of instruction by learning a set of pedagogical principles that can guide their selection or adaptation of materials. This may be so because, no matter how well designed the materials, teachers may not be able to discern principles of their design without explicit guidance. As a consequence, their adaptations may result in enactments that are inconsistent with those principles.

Study Limitations

In this study, our dependent variable was a measure of student learning aligned to district standards. Incorporation of analyses of state assessment data would have yielded more support for our conclusions from the perspective of policymakers and researchers interested in impacts on such measures. In science, however, students are not tested every year, and many tests include too few items in a particular subject area to yield reliable estimates of student achievement in that domain. That was the case in our study, and so we did not analyze impacts of the programs on a state achievement test.

There are also limitations related to the mediation analysis. Mediation analyses intended to identify the "active ingredients" of professional development designs provide evidence of association between teaching and learning outcomes, but they do not provide causal evidence of impact. At best, some view mediational analyses as useful only in generating hypotheses that can guide future research studies because omitted variables may be causal factors and mediators are always measured with error (see, e.g., Green, Ha, & Bullock, 2010, for a critique of mediation models). In the case of the current study, a significant source of error may be due to the fact that we were able to conduct only one observation in teachers' classrooms when units were being enacted (Shavelson, Webb, & Burstein, 1986) and that the measure of assessment strategies did not explore frequency but offered only simple yes–no response options for teachers for each practice. Depending on implementation support available to teachers in an efficacy trial, teachers in a different experimental context might have exhibited greater or lesser variability in their adaptation of the UbD instructional strategies, which could yield different estimates of the mediation effect. Hence, to understand precisely the causal impact of mediators, it is necessary to design a separate experiment to test their impacts and employ a larger number of observations.

Implications for Policy

Scholars studying the history of science education reforms note that policies regarding the roles that teachers should play in designing instruction for students have been inconsistent, with some policies advocating more significant roles for teachers and others suggesting that subject matter experts should design instruction (Atkin & Black, 2003). Evidence from this study supports the idea that consistent and coherent policymaking in science education could target neither teacher skill nor program quality alone but focus instead on supporting teachers' principled, selective use and adaptation of teaching strategies and materials to improve science learning (Cochran-Smith, 2003; Cohen, Raudenbush, & Ball, 2003; Penuel & Gallagher, 2009). Such a policy considers neither teachers nor curricula in and of themselves as agents of change; rather, responsibility for improving students' opportunities to learn depends on "teachers-acting-with" models of good teaching and curriculum materials (M. W. Brown & Edelson, 2003; Remillard, 2005). Reframing the definition of teaching quality in this way, a key aim of educational research becomes to identify programs and models that can develop the knowledge and skills required for teachers to apply principles for selecting and adapting models of teaching and curricular materials for use with their students.

Policies promoting professional development for science teachers fit within this framework, but policies consistent with the framework suggest that the professional development should provide explicit guidance in models of teaching embedded in programs and curriculum. The research does not provide evidence to support the idea that design without guidance can be effective since all of the teachers in the design-oriented programs received guidance in how to design their units. Furthermore, this feature of professional development does not supplant prior research on features of effective professional development (Desimone, 2009); rather, it is an additional critical feature with potential to affect both teachers' instruction and student learning.

By no means is this the first study that argues that embedding models of teaching in professional development is critical for student learning (see, especially Ball & Cohen, 1999; Cohen & Hill, 2001); the fact that our study adds to these calls by presenting evidence in a new domain (Earth systems science) and shows the link between the feature and student learning suggests its broad applicability for guiding policymaking. Additional studies of the approach are still needed to replicate these findings with other widely used models of curricular design in science, including the 5E and project-based learning models, in other domains and at the elementary and high school levels. Finally, research is needed to investigate ways to strengthen models to make their principles more visible to teachers and usable in real classrooms.

Notes

William R. Penuel and Lawrence P. Gallagher are coequal first authors.

[1]TERC is the full name of the organization that developed *ESBD*. TERC used to stand for Technical Education Research Centers.

[2]Differences in the distribution of gender, ethnicity, and grade level were all tested using Fisher's exact test as implemented in Stata Version 11. Differences in teacher years of experience were tested with a two-level hierarchical linear model (teachers nested within schools). The χ^2 result is based on the Wald test of equivalence among the four coefficients representing experimental conditions.

[3]Posttest responses represent student and item performance when students have had an opportunity to learn the material being tested. Thus, they provide more valid data for estimating characteristics of the test items than would responses to the pretest items, when students may have had little knowledge of the material.

[4]One reason why we did not find this association may have been because we did not also consider the cognitive demand of tasks assigned to students to help them learn particular topics.

References

Alozie, N. M., Moje, E. B., & Krajcik, J. S. (2010). An analysis of the supports and constraints for scientific discussion in high school project-based science. *Science Education*, *94*, 395–427.

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

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession* (pp. 3–31). San Francisco, CA: Jossey-Bass.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: Lessons from research on problem and project-based learning. *Journal of Learning Sciences*, *7*, 271–312.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163.

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

Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. *Journal of Research on Educational Effectiveness*, *1*, 237–264.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (vol. *24*, pp. 61–100). Washington, DC: American Educational Research Association.

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

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. (2009). Toward a theory of curriculum design and use: Understanding the teacher-tool relationship. In J. T. Remillard, B. A. Herbel-Eisenman, & G. M. 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.

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

Bybee, R. W. (2004). Scientific inquiry and science teaching. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science* (pp. 1–14). New York, NY: Springer.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins, effectiveness, and applications*. Colorado Springs, CO: Biological Sciences Curriculum Study.

Cochran-Smith, M. (2003). Teaching quality matters. *Journal of Teacher Education*, *54*, 95–98.

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

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 1–24.

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* (vol. *5*, pp. 94–122). New York, NY: Teachers College Press.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Enyedy, N., & Goldberg, J. (2004). Inquiry in interaction: How adaptations of curricula shape classroom communities. *Journal of Research in Science Teaching*, *41*, 905–935.

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.

Geier, R., Blumenfeld, P., Marx, R. W., Krajcik, J., Fishman, B. J., & Soloway, E. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, *45*, 922–939.

Gray, J. T., Camp, P. J., Holbrook, J., Owensby, J., Hyser, S., & Kolodner, J. L. (2001). *Learning by Design technical report: Results of performance assessments for 1999–2000 and 2000–2001*. Atlanta: Georgia Institute of Technology, College of Computing.

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Science*, *628*, 200–208.

Institute of Education Sciences. (2009). *CFDA 84.305A education research grants*. Washington, DC: U.S. Department of Education.

Kennedy, M. M. (1999). Form and substance in mathematics and science professional development. *National Institute for Science Education Brief*, *3*(2), 1–7.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching*, *39*, 522–549.

Krajcik, J. S., & Blumenfeld, P. C. (2006). Project-based learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 317–333). New York, NY: Cambridge University Press.

Krajcik, J. S., & Czerniak, C. M. (2007). *Teaching science in elementary and middle school: A project-based approach* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Krajcik, J. S., Marx, R. W., Blumenfeld, P. C., Soloway, E., Fishman, B. J., & Middleton, M. (2000, April). *Inquiry based science supported by technology: Achievement and motivation among urban middle school students*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Krajcik, J. S., McNeill, K. L., & Reiser, B. J. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education*, *92*, 1–32.

Lin, H.-T., & Fishman, B. J. (2006). Exploring the relationship between teachers' curriculum enactment experience and their understanding of underlying unit structures. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 432–438). Mahwah, NJ: Lawrence Erlbaum.

Liu, X., Congdon, R., & Raudenbush, S. W. (2001). Optimal design for multi-level and longitudinal research (Version 0.23) [Computer software]. Ann Arbor: University of Michigan, Survey Research Center of the Institute of Social Research.

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

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*, 233–268.

McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, *94*, 203–229.

McTighe, J., & Seif, E. (2003). *A summary of underlying theory and research base for understanding by design*. Unpublished manuscript.

McTighe, J., & Wiggins, G. (1999). *The Understanding by Design handbook*. Alexandria, VA: Association for Supervision and Curriculum Development.

McWilliams, H., McAuliffe, C., Lockwood, J., Larsen, J., Rosenberg, M., & Hoover, M. (2006). *Earth Science by Design: A handbook for professional developers*. Cambridge, MA: TERC.

Minstrell, J., & Kraus, P. (2005). Guided inquiry in the science classroom. In National Research Council (Ed.), *How students learn: History, mathematics, and science in the classroom* (pp. 475–514). Washington, DC: National Academies Press.

National Research Council. (1999). *How people learn: Brain, mind, experience*. Washington, DC: National Academies Press.

National Research Council. (2006). *Systems for state science assessment*. Washington, DC: National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

National Research Council. (2010). *A framework for science education: Preliminary public draft*. Washington, DC: Committee on Conceptual Framework for New Science Education Standards, Board on Science Education, National Research Council.

Penuel, W. R., & Gallagher, L. P. (2009). Comparing three approaches to preparing teachers to teach for deep understanding in Earth science: Short-term impacts on teachers and teaching practice. *Journal of the Learning Sciences*, *18*, 461–508.

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: Lawrence Erlbaum.

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263–305). Mahwah, NJ: Lawrence Erlbaum.

Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *Elementary School Journal*, *100*, 331–350.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*, 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: Lawrence Erlbaum.

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

Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching*, *42*, 283–312.

Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teaching. In M. Whittrock (Ed.), *Handbook of research on teaching* (pp. 50–91). New York, NY: 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.

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.

Tabak, I., & Reiser, B. J. (1997). Complementary roles of software-based scaffolding and teacher-student interactions in inquiry learning. In R. Hall, N. Miyake & N.N. Enye (Eds.), *Proceedings of the 2nd International Conference on Computer Supported Cooperative Learning* (pp. 292–301). Toronto, Ontario: International Society of the Learning Sciences.

Thissen, D., Chen, W.-H., & Bock, D. (2003). MULTILOG for Windows (Version 7.0.2327.3) [Computer software]. Lincolnwood, IL: Scientific Software International.

Wartofsky, M. (1979). *Models: Representation and the scientific understanding* (*Vol. 4*). Dordrecht, Netherlands: Riedl.

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*(8), 479–479.

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

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

Manuscript received May 10, 2010
Final revision received March 25, 2011
Accepted April 10, 2011