

Learning-Goals-Driven Design Model: Developing Curriculum Materials That Align With National Standards and Incorporate Project-Based Pedagogy

JOSEPH KRAJCIK

*Center for Highly Interactive Classrooms, Curricula & Computing in Education,
School of Education, University of Michigan, Ann Arbor, MI 48109-1259, USA*

KATHERINE L. McNEILL

Lynch School of Education, Boston College, Chestnut Hill, MA 02467, USA

BRIAN J. REISER

Learning Sciences, SESP, Northwestern University, Evanston, IL 60208, USA

Received 5 January 2007; revised 1 August 2007, 4 August 2007; accepted 11 August 2007

DOI 10.1002/sce.20240

Published online 13 November 2007 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: Reform efforts in science education emphasize the importance of rigorous treatment of science standards and use of innovative pedagogical approaches to make science more meaningful and successful. In this paper, we present a learning-goals-driven design model for developing curriculum materials, which combines national standards and a project-based pedagogical approach. We describe our design model in the context of the Investigating and Questioning our World through Science and Technology (IQWST) project, which is developing a three-year coordinated series of middle grades science curriculum materials. From using this model in the development and enactment of the curriculum, we

This research was conducted as part of the Investigating and Questioning our World through Science and Technology (IQWST) project and the Center for Curriculum Materials in Science (CCMS).

Authorship is in alphabetical order with all authors contributing equally to the conceptualization of the paper.

Correspondence to: Joseph Krajcik; e-mail: krajcik@umich.edu

Contract grant sponsor: National Science Foundation.

Contract grant numbers: ESI 0101780 and ESI 0439493 to the IQWST project at The University of Michigan and Northwestern University, respectively, and ESI 0227557 to the American Association for the Advancement of Science (AAAS) for the CCMS.

Any opinions expressed in this work are those of the authors and do not necessarily represent either those of the funding agency or the University of Michigan, Boston College, or Northwestern University.

identified three important characteristics: unpacking national science standards, developing a learning performances approach to specifying learning goals, and aligning learning goals, instructional activities, and assessments. Using a case study of an IQWST unit from initial development through two cycles of enactment, we describe how these three characteristics help guide curriculum design, identify design issues in curriculum enactments, and guide the development of design solutions. The iterative use of the learning-goals-driven design model coupled with the analysis of multiple data sources informed our revision of the curriculum materials, resulting in substantial student learning gains for the targeted science content and scientific inquiry learning goals. © 2007 Wiley Periodicals, Inc. *Sci Ed* 92:1–32, 2008

INTRODUCTION

Reform efforts in science education call for new instructional materials to improve science teaching and learning. Two aspects of reform efforts are key to improving science curriculum materials: (1) rigorous treatment of science-learning goals (as represented by local, state, and national standards) and (2) use of innovative pedagogical approaches (such as project-based pedagogy) to make science learning more meaningful and support learners in authentic scientific practices. Attention to each of these, standards and innovative pedagogy, carries its own design and implementation challenges, and attempting to achieve them together creates additional complexities. In this paper, we describe a curriculum research and development project devoted to bringing together these goals. We consider the challenges that emerge in pursuing these goals and the design approaches we have developed to address these challenges.

Addressing Content Standards

Clarifying and revising instructional goals for science education has become a central part of reform in the United States and internationally (Black & Atkin, 1996). This has resulted in the development of explicit content standards in the United States (American Association for the Advancement of the Science [AAAS], 1993; National Research Council [NRC], 1996), in standards or curriculum documents specifying reforms in instructional goals in other nations such as England (Millar & Osborne, 1998), Israel (Tomorrow 98, 1992), and Germany (Ertl, 2006), and in international assessments that specify learning goals (Fensham & Harlen, 1999; Organisation for Economic Cooperation and Development [OECD], 2000).

Articulating content standards is central in U.S. systemic reform efforts in science and mathematics (Knapp, 1997; McLaughlin & Shepard, 1995; Smith & O'Day, 1991; Spillane, 2004). There are several arguments for articulating explicit learning goals for science. Accountability policies require agreement about what science is important to teach and therefore to assess (Smith & O'Day, 1991). Standards and curriculum frameworks have been established to ensure sufficient attention to science at particular grade levels (NRC, 1996; Tomorrow 98, 1992), and to establish a vision for what kind of science should be taught (Millar & Osborne, 1998; OECD, 2000; Tomorrow 98, 1992). Content standards enable curriculum guidelines to go beyond specifying a list of topics, such as energy, motion, and ecosystems, and articulate exactly what about these ideas are important to learn (AAAS, 1993).

The success of standards and curriculum frameworks rests on instructional materials aligned with these goals. Thus, current U.S. reform efforts in science education strive to align instructional materials and assessments with local, state, and national standards

(Knapp, 1997; Wilson & Berenthal, 2006). However, few curriculum materials succeed in meeting standards. Project 2061's review of middle school curriculum materials concluded that none of the nine middle school programs they examined would help students learn standards (Kesidou & Roseman, 2002). They criticized materials for covering many topics superficially, and for overemphasis of technical vocabulary. Similarly, analyses of state and local district standards (which drive commercial curriculum developers) have been criticized for shallow coverage of many topics (Schmidt, Wang, & McKnight, 2005). The Project 2061 analyses also found that curriculum materials did not build on student-learning research (Kesidou & Roseman, 2002) as materials failed to take into account students' prior knowledge, lacked coherent explanations of phenomena, and failed to support students in developing explanations of phenomena.

Working with standards poses challenges for design of curriculum materials. Although standards provide guidelines as to which aspects of science to address, they are statements of scientific ideas and skills from the perspective of science, and organized according to how experts view relationships between ideas. Using standards as guides for instruction requires designers to go further to consider four important facets of design: (1) how to make these ideas compelling and understandable to learners, (2) what a psychological or learning-based account of these ideas would entail, (3) what kinds of experiences would help learners develop these ideas, and (4) what kinds of reasoning tasks would represent the use of this knowledge.

Project-Based Pedagogy

A second aspect of reform involves a shift in pedagogy to make science learning meaningful and more focused on *learning* science by *doing* science. Several important pedagogical themes emerge from several countries (Black & Atkin, 1996). A focus on learning through investigations is key to reforms (Abd-El-Khalick et al., 2004; Millar & Osborne, 1998; Tomorrow 98, 1992). Participating in the practice of science as well as learning how science functions as a discipline are two aspects of the move toward scientific investigations (Duschl, Schweingruber, & Shouse, 2007; McComas & Olson, 1998). Understanding how scientists build, evaluate, and apply scientific knowledge is a core part of this emerging consensus view of scientific literacy (AAAS, 1993; Fensham & Harlen, 1999; OECD, 2000). Another shared aspect is an increased focus on connecting science understanding to learners' experiences with the everyday world (Millar & Osborne, 1998; OECD, 2000; Tomorrow 98, 1992).

Research on project-based science (PBS) has explored approaches to address these goals, by embedding science learning in investigations of meaningful real-world problems (Blumenfeld & Krajcik, 2006; Edelson, 2001; Linn, Bell, & Davis, 2004; Reiser et al., 2001; Zohar & Nemet, 2002). These approaches apply the basic tenant of cognitive apprenticeship in which learners apply scientific ideas and skills to investigate and solve meaningful problems (Collins, Brown, & Newman, 1989). PBS provides a meaningful context to involve learners in the practices of knowledge building (Duschl et al., 2007; Lehrer & Schauble, 2006). These approaches involve learners in scientific practices such as argumentation, explanation, scientific modeling, and engineering design (Bell & Linn, 2000; Fretz et al., 2002; Kolodner et al., 2003; McNeill, Lizotte, Krajcik, & Marx, 2006; Osborne, Erduran, & Simon, 2004; Sandoval & Reiser, 2004; Zohar & Nemet, 2002). PBS also helps students develop explanations of scientific phenomena that represent important disciplinary understandings (McNeill & Krajcik, 2007; Reiser et al., 2001; Sandoval, 2003) and understand core scientific ideas as represented by science standards (Marx et al., 2004; Rivet & Krajcik, 2004). Rich application problems that situate the scientific ideas and skills

can build connections between students' scientific knowledge and their understanding of everyday experiences (Linn et al., 2004).

Design Tensions Between the Standards and Project-Based Pedagogy

The ideas of project-based pedagogy are consistent with the intent of science standards, in a broad strokes analysis. However, tensions arise when designing materials that carefully treat science standards and use a project-based pedagogy. One potential design tension arises between content choices dictated by the problem context versus the standards. A core tenet of PBS is examining ideas in depth. Although problems such as global warming, pollution, or water quality can serve as vehicles for uncovering the need for scientific knowledge, investigating these problems may go beyond knowledge stated in standards. Similarly, a tension may emerge in sequencing science ideas according to content connections versus the order in which ideas arise in addressing real-world problems (Sherin, Edelson, & Brown, 2004).

A second potential design tension exists between the depth of coverage of content in PBS and the needs of addressing a full year's curriculum of standards. To date, efforts at project-based pedagogy have primarily explored a "replacement unit" approach, in which units are integrated into teachers' ongoing curricula. For example, the Learning By Design units (Kolodner et al., 2003), units developed in the Center for Learning Technologies in Urban Schools (Singer, Marx, Krajcik, & Chambers, 2000), and units designed in response to Israel's Tomorrow 98 initiative (Tal & Kedmi, 2006) are generally 6–8 weeks in length and modular in design. The Web-based Integrated Science Education (WISE) investigations are modular week-long units to maximize their potential for integration into existing curricula in various settings (Linn et al., 2004). The conflict between in-depth treatment of scientific ideas in PBS and the breadth of coverage demanded by many local standards is a very real tension teachers perceive (Aikenhead, 2005; Schneider, Krajcik, & Blumenfeld, 2005). Developing full-year project-based materials may require more careful treatment of content standards to ensure breadth of coverage than in prior replacement-unit approaches.

A third potential design tension arises from meeting standards within the contextualized nature of project-based pedagogy. PBS often uses problem scenarios to motivate and make the science meaningful (Edelson, 2001; Sherin et al., 2004). Yet, the focus on specific contexts creates an additional need to address the generality articulated in standards. Mastery of the specific ideas in the project context is not sufficient. For example, while teaching form and function in animals might be easily integrated into a problem-based unit on change in an ecosystem resulting from introduced predators, generalizing the idea to plants (as stated in the standard) requires additional attention in the design. A potential pitfall in PBS is that learners may focus unduly on solving the problem or "engineering" a desired state of affairs, rather than working toward generalizations about why a solution works (Barron et al., 1998; Schauble, Klopfer, & Raghavan, 1991).

In this paper, we describe our design process for bringing together rigorous treatment of standards and use of project-based pedagogy. We are investigating this design approach in the *Investigating and Questioning our World through Science and Technology* (IQWST) project, a research-based 3-year coordinated series of middle grades science curriculum materials. We illustrate how our design process can result in materials that achieve these dual goals of pedagogy and standards to promote learning. We describe the design challenges that arise in pursuing these dual goals, and how our design process helps uncover and address these issues.

We describe our *learning-goals-driven* design model that builds upon and extends the backward design approach presented by Wiggins and McTighe (1998) and current

instructional design frameworks (Gagné, Wager, Golas, & Keller, 2005). We present the model using examples drawn from a project-based seventh-grade chemistry IQWST unit. We review three important characteristics of the design model that enable us to combine project-based pedagogy and careful treatment of standards: unpacking standards from a learning perspective, a learning performances approach to specifying learning goals, and aligning learning goals, activities, and assessments. In the empirical section of the paper, we describe how we used these design characteristics to guide the analysis of a first enactment, to identify design issues to be addressed, and to guide treating these design issues in curriculum revisions.

LEARNING-GOALS-DRIVEN DESIGN MODEL

In this section, we describe the learning-goals-driven design model and examine how the different aspects of the model worked together to allow designers to create materials that blend content standards with project-based pedagogy. To illustrate the design model, we provide examples from one of the first curriculum units we developed in IQWST, *How can I make new stuff from old stuff?* (Stuff) (McNeill et al., 2004). This unit focuses on three central ideas in chemistry—the conservation of matter, substances and their properties, and chemical reactions.

The learning-goals-driven design model includes three stages: (1) specifying learning goals, (2) developing materials, and (3) gathering feedback. Figure 1 illustrates the component processes of these different stages and their iterative nature essential for aligning learning goals, materials, and assessments. In the next sections, we describe each stage and how we address content standards and project-based pedagogy.

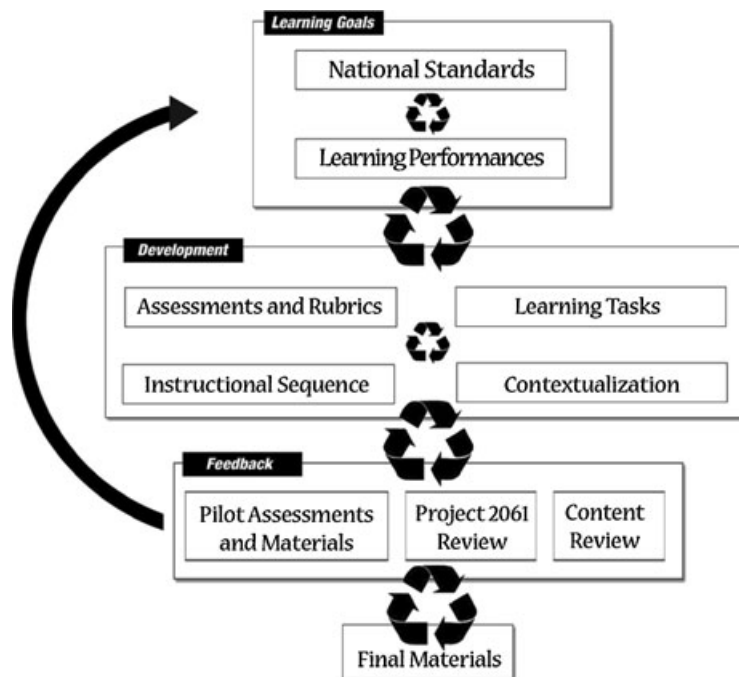


Figure 1. Learning-goals-driven design model.

Moving from Content Standards to Learning Goals

Identifying and Unpacking Standards. To rigorously address the national standards, we spend considerable time identifying and unpacking appropriate science standards. We identify key content from the national science standards, *Benchmarks for Science Literacy* (AAAS, 1993), the companion document, *Atlas of Science Literacy* (AAAS, 2001), and the *National Science Education Standards* (NRC, 1996). We create maps that include a coherent cluster of content standards as well as requisite prior knowledge and common misconceptions for each instructional unit. (See Appendix A for a map for the chemistry ideas in the Stuff unit.) Creating these maps is essential because it allows us to identify clusters of standards that interrelate, determine what are the big ideas in the standards, and build an instructional sequence to foster more complex understandings over time. These maps are similar to and informed by the maps in the *Atlas of Science Literacy* (AAAS, 2001). Yet they are different in that we created them from a pedagogical perspective with the goal of developing curriculum units. The standards that end up in a map are influenced by project-based pedagogy, such as the need to connect to students' own experiences and contextualize the units in real problems. Consequently, our maps end up somewhat different than the AAAS maps that focus solely on connections between concepts. For example, our maps include a more focused subset of standards with some different connections among them. Creating maps of content standards was not a part of our design process in our previous project-based work to develop replacement units. However, mapping these relationships among ideas both within and across units is essential for a curriculum sequence that develops coherently over time and that addresses the big ideas in science.

Next, we “unpack” each standard, breaking apart and expanding the various concepts to elaborate the intended science content. This unpacking is guided simultaneously by fidelity to the scientific ideas articulated in the standard and by consideration of learning and pedagogical concerns. When we unpack a standard, we consider what content is important, as well as what aspects are suitable for middle school students by examining common student difficulties, prior understandings needed to build the target understanding, and aspects of prior conceptions that may pose challenges. Appendix B provides the unpacking of the five standards in the Stuff unit. For example, when we considered the implicit complexity of ideas in the chemical reaction standard that might not be transparent to learners, we found this standard required a scientific understanding of the ideas of “substance” and “property.” Therefore, we added a standard (from NRC, 1996) to represent these learning goals. Although this new standard was not one of our initial learning goals, unpacking the elements necessary for reasoning about these ideas uncovered properties and substances as ideas the unit would need to address.

Another aspect of our learning-based unpacking is to help designers focus on what are the essential parts of the scientific ideas. For example, in an IQWST unit on ecosystems, *Struggle in natural environments: What will survive?* (Bruozas et al., 2004), we unpacked the standard on competition between species, which states “In all environments, freshwater, marine, forest, desert, grassland, mountain, and others, organisms with similar needs may compete with one another for resources including food, space, water, air and shelter” (AAAS, 5D1). When we mapped the relationships between learning goals, it was clear that we need not exhaustively treat all six types of environments and all types of resources mentioned in the standard to teach the target idea of competition for limited resources. In contrast, in unpacking other standards dealing with food for plants and animals, it was clear that the commonalities and differences of the understanding about food for plants and animals involved understanding different kinds of mechanisms, and each needed to be treated in depth. Thus, to both uncover understandings that may be implicit, and help focus on what

understandings to treat in depth, treating standards rigorously entails considering the science concepts articulated in the standards in depth and how they link to other science ideas.

Developing Learning Performances. Once we identify and unpack the key science standards, we develop “learning performances” to articulate the cognitive tasks for students to accomplish. The development of learning performances is a critical step. Science standards are declarative statements of scientific ideas, and as such do not specify the type of reasoning we would like students to engage in with these concepts. An important theme in science reform has been a focus on meaningful understanding, in which the goal of learning is for students to be able to reason with scientific knowledge. Views of scientific knowledge have shifted from a collection of concepts and skills to a “knowledge in use” view of science as a knowledge-building practice, in which people construct explanatory models and use them to make sense of data, make predictions, develop explanations, and refine their models (Duschl et al., 2007; Lehrer & Schauble, 2006). Thus, science understanding is best articulated as learning goals by specifying the kinds of reasoning tasks that use that knowledge in sensemaking. Building on Perkin’s (1998) notion of “understanding performances,” we call the specification of scientific reasoning with particular science content “learning performances.” The learning performances serve as the learning goals that guide development of learning activities and assessments.

To develop learning performances, we draw on the scientific practices described by the habits of mind standards (AAAS, 1993) and the scientific inquiry standards (NRC, 1996) to consider what types of cognitive tasks we want students to accomplish. We also adapt the revised Bloom’s taxonomy (Anderson & Krathwohl, 2001) to focus on scientific disciplinary thinking to characterize cognitive tasks that students might perform. We identify a variety of practices for students to engage in such as designing investigations and creating models. Table 1 illustrates the process of developing learning performances by crossing a content standard with a practice. In this case, students use their understanding of chemical reactions to write a scientific explanation about a phenomenon in which they argue that either a chemical reaction did or did not occur and then justify that claim with appropriate evidence and reasoning.

Creating learning performances combines rigorous treatment of science standards with project-based pedagogy. Engaging in this process requires serious consideration of the

TABLE 1
Developing Learning Performances

Content Standard	× Practice (Scientific Inquiry Standard)	= Learning Performance
When substances interact to form new substances, the elements composing them combine in new ways. In such recombinations, the properties of the new combinations may be very different from those of the old. (AAAS, 1990, p. 47)	Develop . . . explanations . . . using evidence. (NRC, 1996, A: 1/4, 5–8) Think critically and logically to make the relationships between evidence and explanation. (NRC, 1996, A: 1/5, 5–8)	Students construct scientific explanations stating a claim whether a chemical reaction occurred, evidence in the form of properties, and reasoning that a chemical reaction is a process in which old substances interact to form new substances with different properties than the old substances.

content and inquiry standards and the relationship between the two. This approach to articulating learning goals goes beyond “understanding” as a learning goal to specify how the knowledge is used, and specify an integrated practice of using scientific ideas to explain the world (Duschl et al., 2007), in contrast to what in the past has been specified as separate content and process goals. We conceptualize the scientific understandings from the start as embedded in practice (e.g., explaining why a chemical reaction occurred, predicting population changes), rather than as static facts. This approach to constructing performances enables us to apply the same aspect of scientific practices (modeling, argumentation, explanation) across different scientific problems and disciplines, resulting in a consistent treatment of scientific practices across chemistry, physics, earth science, and biology content areas. Developing learning performances is also an important aspect of achieving project-based pedagogy, specifying how knowledge is to be applied (Collins et al., 1989; Edelson, 2001).

Development Stage

The development stage is guided by attention to alignment throughout the curriculum and assessment materials. In this stage, we develop the contextualization, learning tasks, instructional sequence, assessments, and rubrics.

Contextualization. The contextualization step in design connects the treatment of learning goals to students’ own experiences and real-world problems. One way we contextualize units is with a driving question, a rich and open-ended question in everyday language that draws on interests and curiosities students have about the world (Blumenfeld & Krajcik, 2006; Blumenfeld et al., 1991). The driving question articulates a problem context that creates the need for the scientific understandings (Edelson, 2001). Students gain understanding of the target learning goals as they attempt to address the driving question. In developing the Stuff unit, we struggled to contextualize the unit in a question that might interest students and connect to their lives outside of school, while simultaneously allowing us to address the chemistry standards. With input from teachers, we developed the driving question “How can I make new stuff from old stuff?” This driving question met four criteria for driving questions: the question should be worthwhile, feasible, grounded in real-world problems, and meaningful (Krajcik & Czerniak, 2007). Students explore this question by making new substances from old substances, specifically making soap from fat (lard) and sodium hydroxide. We use students’ everyday experiences with fat and soap as well as other phenomena to help them see that chemical reactions do not only happen in science classrooms, but rather occur constantly in the everyday world.

Learning Tasks. After unpacking the standards and developing learning performances, we create instructional tasks designed to help students develop mastery of the learning goals. When developing tasks, we strive to find phenomena that align with the learning goal, make complex scientific ideas plausible for students, and help students appreciate the utility of scientific concepts (Kesidou & Roseman, 2002). In the Stuff unit, we contextualized the learning goals in the challenge of making soap from fat. We also identified a number of other appropriate phenomena for the chemical reaction learning goal such as burning magnesium (which occurs in sparklers and fireworks), and reacting a copper penny with vinegar (similar to the chemical reaction that created the green substance, copper sulfate, on the surface of the Statue of Liberty). These phenomena align with the learning goal, are feasible in a middle school classroom setting, and connect to students’ everyday experiences. Once we

identify candidate phenomena, the learning performances guide design of learning tasks by specifying cognitive tasks we want students to do with the phenomenon. For example, in the learning tasks concerning the chemical reaction in which students make soap from fat, students collected data from the phenomenon and constructed a scientific explanation to justify whether a chemical reaction occurred.

Instructional Sequence. Next, we create a coherent instructional sequence to help build student understanding and answer the driving question. We return to our map of learning goals to consider what science concepts should come first to build more complex understandings. We also consider what knowledge is necessary and how we will create a need for that knowledge so that the task can be incorporated into an investigation of the driving question. In the Stuff unit, we first focused on the concept of substances and properties and then on chemical reactions. The instructional sequence contained a number of investigations that students completed, allowing them to cycle back to these ideas of substances and properties as well as build on them to develop an understanding of chemical reactions.

Assessments. We develop assessment items that are aligned with the learning performances, which characterize the cognitive performances that can be assessed. We describe how we develop the assessment items and the corresponding rubrics elsewhere (Harris, McNeill, Lizotte, Marx, & Krajcik, 2006; McNeill & Krajcik, 2007). The key point here is that the curriculum materials and assessments are designed together in an iterative process that align with both learning performances and the science standards. Using the learning performances to guide assessments ensures that both our goals and evidence for learning are defined in terms of integrated use of scientific ideas. Looking at practices across different content in the assessment tasks provides us with a richer measure of students' ability to engage in that practice.

Feedback

The last stage of the learning-goals-driven design model uses feedback from a variety of sources to revise the curriculum materials. We gather feedback and assessment on different aspects of the effectiveness of curriculum materials. Researchers from Project 2061 analyze the curriculum materials (Kesidou & Roseman, 2002), examining alignment with learning goals and instructional supports. Experts in the scientific discipline conduct content reviews for scientific accuracy. We gather empirical data from classroom enactments of teachers and students using the materials. We collect fieldnotes, videotapes, and students' artifacts to investigate the alignment between the designed learning performances and the teaching-learning interactions that occur in classrooms. We administer pre- and posttest measures of student learning. Finally, we gather teacher feedback through informal interviews during and following enactments.

Contrast With Earlier Instructional Design Approaches

The learning-goals-driven design model builds on earlier ideas of instructional design in several respects. Like backwards design (Wiggins & McTighe, 1998), we stress the importance of specifying clear objectives for what learners will be able to do after instruction, echoing a common prescription in instructional design approaches (Gagné et al., 2005; Tyler, Gagné, & Scriven, 1967). The process of specifying goals, unpacking, designing tasks, and iterating with feedback mirrors more general models of instructional design (ISD) (Gagné et al., 2005). For example, Gagné and his colleagues presented a general

framework for instructional design consisting of the sequence of stages: analyze, design, develop, implement, and evaluate.

These instructional design approaches provide a general non-discipline-specific framework to guide design. In contrast, we present a science-specific model that focuses on disciplinary challenges such as the unpacking of science standards and operationalizing scientific reasoning as learning performances. The learning-goals-driven design model refines the ISD “analyzing objectives” through a learning-based approach to selecting and unpacking learning goals guided by research on science learning (Duschl et al., 2007). As we described, selecting and unpacking learning goals involves examining standards to identify core ideas, analyzing the conceptual structure of the target ideas to decompose complex ideas, uncovering tacit understandings needed to understand the ideas, and considering students’ prior conceptions that present opportunities and challenges in building these ideas. A related difference is in articulating learning performances, which like ISD “task analysis” specify the task, but are developed according to research-based models of cognitive performances in science. In summary, learning-goals-driven design is similar in spirit to these earlier design frameworks, but builds on theories of conceptual understanding and cognitive performance to create a disciplinary-learning approach to design. In the next section, we describe how the learning-goals-driven design model guides our collection and use of data in evaluation and revision.

GUIDING ANALYSIS AND REDESIGN

In developing IQWST materials, we engaged in multiple cycles of the learning-goals-driven design model, with each cycle informed by theoretical and empirical feedback. To study the effect of the curriculum materials, it is important to study how teachers and students use the materials in actual classroom contexts. The complexity of the classroom environment poses challenges for the usability and effectiveness of the materials. Consequently, our multiple data sources came from both external reviewers examining our materials, and our own design-based research investigating the materials as enacted by teachers and students (Brown, 1992; Edelson, 2002). Design-based research combines the design and implementation of innovative-learning environments with the simultaneous systematic study of those innovations within the context for which they were created (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; The Design-Based Research Collective, 2003). The core ideas of unpacking, learning performances, and alignment help focus what data to collect, the analysis process, and the use of the feedback in revising the curriculum. In this section, we examine the kinds of design issues that emerge in developing curriculum materials that rigorously address standards and incorporate project-based pedagogy, and describe how the learning-goals-driven design model helps identify design challenges and address those challenges in the revision process.

Participants and Setting

During the first enactment in 2001–2002, two teachers implemented the Stuff unit in public middle schools. The teachers taught in different urban areas, with a total of 119 seventh-grade students (see Table 2). The two urban sites, Urban A and Urban B, are large cities in the midwest. The students in the school in Urban A were over 90% African American and came from lower to lower-middle income families. The students in the school in Urban B consisted of a more ethnically diverse population (approximately 51% African American, 45% Caucasian, 3% Hispanic, and 1% Asian) and came from lower-middle income families.

TABLE 2
Teachers, Students, and Classrooms Involved in the First Enactment

Site	Urban A	Urban B	Total
Schools	1	1	2
Teachers	1	1	2
Classrooms	1	3	4
Students	31	88	119

Data Sources

We collected a variety of data sources that allowed us to evaluate the curriculum materials in terms of their learning goals alignment and pedagogical supports, and the enactments of the curriculum in terms of classroom focus on student learning, engagement in scientific practices, and participation in project-based pedagogy. The data sources included the following: student pre- and posttests, student artifacts, fieldnotes, selected classroom videos, teacher feedback, Project 2061 review, and content expert feedback. We used these data sources to identify and triangulate design issues to be addressed in subsequent curriculum revisions.

Identical pre- and posttest measures consisted of 20 multiple-choice and 4 open-ended items that targeted the key chemistry ideas (see Appendix A) and the learning performances developed for the unit (see Appendix B). The assessment items and rubrics are described in more detail elsewhere (Harris et al., 2006; McNeill & Krajcik, 2007). We only included students in the analysis who completed both the pre- and posttest assessments. We scored the multiple-choice responses for a maximum possible score of 20. We developed rubrics to score the four open-ended items with a maximum possible score of 15. Pairs of independent raters scored the open-ended items using specific rubrics with an average interrater reliability of 90%. A third independent rater resolved all disagreements. In the Urban A enactment, we also collected and analyzed student artifacts, particularly focusing on students' written scientific explanations. Similar to the open-ended test items, we used rubrics to score student artifacts. In this case, raters assigned scores through discussion and then came to agreement by consensus. We used the artifacts to test claims about student learning that arose from our analysis of the test data.

We examined classroom fieldnotes from Urban A classrooms looking for general themes in terms of alignment with the standards and project-based pedagogy. On the basis of these themes, we examined selected videotapes from the enactments looking for confirming and disconfirming evidence to support our hypotheses from the fieldnotes.

Teacher feedback provided another data source on the quality and usability of the curriculum materials. Teachers participated in weekly telephone conversations during the implementation of the curriculum in which they provided feedback on classroom implementation. We also held a wrap-up meeting after they completed the unit to obtain their reflections and discuss specific concerns in more depth that arose during the telephone conversations and from other data sources.

Finally, we received feedback on the materials themselves from two external sources. Project 2061 performed a preliminary analysis of alignment and pedagogical supports using the criteria described by Kesidou and Roseman (2002). We also received feedback from a content expert in chemistry who reviewed our treatment of chemistry content.¹

¹ After the second enactment of the unit, we had two content experts with PhDs in chemistry complete a more thorough review of the curriculum materials in which they read the unit in its entirety and evaluated the appropriateness of the content.

TABLE 3
Enactment 1 Test Data for Substances and Properties Versus Chemical Reactions

Site	Pretest <i>M</i> (<i>SD</i>) ^a	Posttest <i>M</i> (<i>SD</i>)	<i>t</i> -Value ^b	Effect Size ^c
Urban A (<i>n</i> = 12)				
Substance and property	4.00 (1.91)	6.92 (2.64)	4.61***	1.53
Chemical reaction	4.13 (2.05)	6.46 (2.48)	2.61*	1.14
Urban B (<i>n</i> = 77)				
Substance and property	6.32 (2.46)	9.27 (3.09)	9.52***	1.20
Chemical reaction	6.07 (2.24)	7.64 (2.20)	6.56***	0.70

^aMaximum score: Substance and property = 15, Chemical reactions = 16.

^bOne-tailed paired *t*-test.

^cEffect Size: Calculated by dividing the difference between posttest and pretest mean scores by the pretest standard deviation.

p* < .05, *p* < .01; ****p* < .001.

Design Issues and Revision of Curriculum

We used our learning-goals-driven design model to identify design issues in meeting standards and supporting project-based learning. As in the materials design, three characteristics of the model guided our analysis: unpacking standards, learning performances, and alignment throughout curriculum and assessments. These aspects provided the lenses to guide evaluation of data to detect problems and informed solutions to these identified issues. We focus our discussion on the feedback and revision for one of the target science standards, chemical reactions. As Table 3 illustrates, students at both sites did not have as large learning gains for the chemical reaction learning goal as they did for the substance-and property learning goal, suggesting revisions were needed. Although we concentrate on the chemical reaction standard here, we used a similar process to identify and address design issues for all learning goals.

Our analysis of the chemical reaction learning goal identified five general design issues, which we suggest could arise when creating any science curriculum. Table 4 presents a summary of each design issue, a specific example from the Stuff unit, data sources, and evidence from the data source that allowed us to identify the issue. We discuss each of these in greater detail to illustrate how we used the learning-goals-driven design model to both identify the design issue and revise the curriculum.

Design Issue #1: Rationale and Alignment of Learning Performances. One initial issue was our use of learning performances as the key learning goals of the unit. Project 2061's review suggested that they did not understand our rationale behind learning performances. Their review stated that our learning performances were not sufficiently aligned with the national standards in that they asked students to "go beyond" what was articulated in the standard (see Table 4).

Although as a research group we had discussed the role of learning performances and why we viewed engaging in scientific practices as a core approach to supporting learning, the Project 2061 review revealed that we had not included this rationale in the curriculum materials. For example, representations like Table 1 demonstrating how we combined a content standard with a practice were not included in the curriculum materials. We did

TABLE 4
General Design Issue, Example, Data Source, and Evidence for Chemical Reaction Revisions

General Design Issue	Example	Data Source	Evidence from Data Source
#1: Rationale and alignment of learning performances	There was not a rationale for the learning performances and they were not sufficiently aligned.	Project 2061	"... there is not a good match between learning goals [standards] and learning performances ... it's important to check the performances against the learning goals to be sure that a) the knowledge in the learning goal is needed for the performance and b) students can carry out the performance with the knowledge in the learning goal." (Review #2, p. 2)
#2: Need to unpack inquiry standards	Curriculum materials did not explicitly state what is meant by "explain" and did not provide teachers or students guidance in creating "Explanations"	Pre- and posttests Student artifacts Project 2061	<p>Showed student difficulty in including the evidence and reasoning components of a scientific explanation (see Table 5)</p> <p>Showed student difficulty in including reasoning component of a scientific explanation.</p> <p>"... it is not clear what is meant by the phrase "explain that" in learning performances 3 and 9." (Review #1, p. 2)</p>
#3: Aligning science ideas across multiple contexts	Example 1: Properties were not integrated into the chemical reaction portion of the unit	Project 2061	"Students spend all this time up front investigating density, solubility, melting point as characteristic properties of substances but then do not use these properties in subsequent lessons to establish whether a chemical reaction has occurred or not. Instead, different properties (thickness/runniness or state of the material) turn up without explicit preparation." (Review #1, p. 7)
	Example 2: Atoms and molecules were not included in the unit	Project 2061	"Elements are not mentioned ... if the idea of a small number of elements is not to be developed until later, what is the rationale for this approach? Four weeks seems a large investment for a small payoff." (Review #1, p. 6)

Continued

TABLE 4
(Continued)

General Design Issue	Example	Data Source	Evidence From Data Source
#4: Students' overgeneralizations of concepts from exemplars	After the unit, students thought mixtures were a chemical reaction	Teacher feedback Pre- and posttests	Introducing atoms and molecules earlier would help students' understandings of substances and chemical reactions. On multiple-choice item 13, more students thought making lemonade was a chemical reaction after the unit than before. On the open-ended items, students included "dissolving" as evidence for a chemical reaction.
#5: Iterative alignment of assessments with learning goals	Pre- and posttest did not adequately align with learning goals	Content expert Project 2061 Pre- and posttests	Discussed with content expert the distinctions between dissolving and chemical reactions as well as the most appropriate way to discuss this with middle school students. "It is of some concern that the Test at the beginning is not well focused on assessing the learning goals or key ideas . . . For nearly all of the items, the key ideas are either not necessary or not sufficient for responding correctly." (Review #1, p. 10) Suggested that some of the questions did not effectively assess the desired learning goal.

not discuss in the curriculum how we viewed the learning performances as a way of operationalizing the science content to specify what we wanted the students to *do* with the science content. To address this issue, we developed an introductory section in the curriculum materials that discussed the rationale and how we developed and used the learning performances.

Furthermore, Project 2061's critique identified that the language in the learning performances was not always consistent or explicit across lessons. The learning performance in Table 1 explicitly states what students should include in their scientific explanation (i.e., claim, evidence, and reasoning). However, many learning performances were less clear, for example, "students will be able to explain whether or not an observable change corresponds to a chemical reaction," thus failing to articulate what was meant by "explain." Consequently, we revised the learning performances to more clearly articulate what we meant by the different practices.

Another key characteristic of our design model is alignment across standards, curriculum, and assessment. Project 2061's critique identified the need to align learning performances more explicitly with the other components. Consequently, we developed a more careful taxonomy of learning performances, deriving them directly from a combination of content and inquiry standards. We decided to focus on eight cognitive practices across IQWST units: defining scientific concepts, identifying examples of concepts, designing investigations, conducting investigations, analyzing and interpreting data, constructing scientific explanations, constructing models to represent processes, and using models to reason about processes. We used these learning performances in our revisions of lessons and assessments to ensure that we were using the same language and practices throughout. Appendix C provides the learning performances for the chemical reaction content and demonstrates how we constructed different learning goals with a range of cognitive demands from the same content standard. We had the students engage in the same eight practices across the three different content areas in the unit (substance and properties, chemical reactions, and conservation of mass) to build a more robust understanding of the practices. For example, we had students write scientific explanations for whether two substances are the same, whether a chemical reaction occurred, and whether mass changed. In summary, expert feedback on alignment helped identify the need to provide a rationale for learning performances as learning goals, include an explicit articulation of these goals, and subject these performances to the same alignment evaluation as content standards.

Design Issue #2: Need to Unpack Inquiry Practices. Project 2061's review of our learning performances specifically articulated a concern that the learning performances had not clearly defined what is meant by "explain" (see Table 4). In our own discussions, we had unpacked scientific explanation to develop an instructional framework in which we broke down the practice into three components: claim, evidence, and reasoning (McNeill et al., 2006; Moje et al., 2004). The *claim* is a conclusion about the question or problem. The *evidence* is scientific data that support the claim. Finally, the *reasoning* is a justification for why the evidence supports the claim, which often includes appropriate scientific principles. Yet in reviewing the first version of the curriculum, we realized this unpacking was not provided as an explicit structure in either the student or teacher materials. While national standards clearly call for students to explain scientifically, there is not an empirically based instructional model that specifies how to support students in this practice. Just as it is necessary to clarify the content goals, we found this essential for inquiry practices. When we analyzed students' pre- and posttest written explanations, we also had concerns around student learning of scientific explanations (Table 5).

TABLE 5
Enactment 1 Data for Scientific Explanation

Site	Pretest <i>M</i> (<i>SD</i>) ^a	Posttest <i>M</i> (<i>SD</i>)	<i>t</i> -Value ^b	Effect Size ^c
Urban A (<i>n</i> = 12)				
Total	1.88 (1.75)	3.19 (1.75)	1.91	0.75
Claim	1.25 (1.31)	2.08 (0.97)	1.77	0.63
Evidence	0.52 (0.78)	1.01 (0.96)	1.77	0.63
Reasoning	0.10 (0.36)	0.10 (0.36)	0.00	0.00
Urban B (<i>n</i> = 77)				
Total	2.85 (1.49)	3.71 (1.25)	4.61***	0.58
Claim	1.95 (1.04)	2.21 (0.81)	1.92	0.25
Evidence	0.74 (0.72)	1.21 (0.83)	4.15***	0.65
Reasoning	0.16 (0.42)	0.29 (0.53)	2.04*	0.31

^aMaximum score: Total = 9, Claim = 3, Evidence = 3, Reasoning = 3.

^bOne-tailed paired *t*-test.

^cEffect Size: Calculated by dividing the difference between posttest and pretest mean scores by the pretest standard deviation.

p* < .05; *p* < .01; ****p* < .001.

Students in Urban A did not achieve significant learning gains for explanation as a whole or for any of the individual components. Although students in Urban B did have significant learning gains for explanations as a whole, students' mean posttest scores for evidence and reasoning at both sites reveal that students had difficulty with evidence and reasoning aspects of constructing scientific explanations.

We then analyzed the explanation students wrote as part of the classroom instruction (see Harris et al., 2006). We found that while students' claim and evidence scores improved, they still frequently failed to include reasoning in their explanations. For example, for one chemical reaction task, we were looking for students to articulate that they knew a chemical reaction had occurred when the new substances formed because they had very different properties from the old substances. Yet students' reasoning rarely justified why their evidence supported the claim by using the underlying scientific principle. For example, a typical students' reasoning stated, "This evidence supports that a chemical reaction occurred because you can follow the evidence and determine that it change."

Our learning-goals-driven design model relies on unpacking of learning goals into their constituent aspects, to ensure that selection of phenomena and design of tasks engage learners with the different aspects of the idea or practice and to support these different aspects of the understanding. The feedback revealed that we had not made the unpacking of inquiry practices transparent in the materials for teachers or students. Making the structure of complex tasks more explicit for students can be a key part of scaffolding that practice (Quintana et al., 2004). To address this issue, we revised our materials to use the unpacking of inquiry practices to provide needed supports. We began by revising the learning performances to state what we wanted students to include in their scientific explanations (i.e., claim, evidence, and reasoning). These new learning performances then guided revision of both the curriculum and assessments. We added a lesson to the unit in which teachers introduced scientific explanation to the students through a variety of instructional strategies including defining scientific explanation and providing a rationale for the practice (McNeill

& Krajcik, in press). Furthermore, we added written curricular scaffolds, which supported students with each of the components (McNeill et al., 2006). Finally, we included test items that assessed explanations across different content standards to give us a more complete picture of students' ability to write scientific explanations. Thus, the revisions of support for scientific explanation drew on all three key characteristics of the learning-goals-driven design model—unpacking the learning goals, developing explicit learning performances to reflect these goals, and aligning tasks and assessments with the learning goals. Unpacking the scientific inquiry practices for teachers and students provides a learner-appropriate model of inquiry, and uncovers where support is needed to engage learners in these practices essential for project-based pedagogy.

Design Issue #3: Aligning Science Ideas Across Multiple Contexts. Project 2061's review suggested two examples for which the materials would be stronger if they reused the key science concepts in multiple contexts and created greater alignment between the learning goals and curriculum. First, they critiqued the lack of integration of properties into the chemical reaction segment of the materials. In the beginning of the unit, students determined melting point, solubility, density, hardness, and color of both fat and soap. While they again investigated these properties for their homemade soap, they did not use melting point, solubility, or density in the interim lessons to reason about chemical reactions. Project 2061 was concerned that without repeated opportunities and greater alignment, students would not integrate their understanding of properties with chemical reactions.

When revising the unit, we were concerned about students determining the density, solubility, and melting point for substances in all experiments, both because of the time to complete these procedures and the difficulty of making some of the measurements (e.g., the density of oxygen is 0.00131 g/cm^3). We attempted to handle the trade-off between pedagogical advantages and logistical disadvantages by revising some chemical reaction experiments so that students collected data on the properties of the substances, whereas for other experiments we provided students with the properties of the substances. For example, for an electrolysis of water experiment, students would not be able to determine the melting point, solubility, and density for hydrogen and oxygen gas, so we provided a table with this information, which students analyzed to determine whether electrolysis was a chemical reaction. These opportunities in the second learning set allowed students to use their developing understanding of properties in a new context and offer greater alignment.

The second example focuses on how we originally had planned to include the particle nature of matter as a final learning set in the unit, not integrated into students' earlier experiences with substances, properties, and chemical reactions. Project 2061 argued that we should introduce the particle nature of matter earlier and use it across the multiple different contexts (see Table 4). In debriefing interviews, pilot teachers agreed that the particle nature of matter and the concepts of molecules could be brought in earlier and might increase students' understanding of substances and chemical reactions.

This issue led us to reconsider how we unpacked these content standards and include a particle-level interpretation in our unpacking of standards earlier in the unit. For example, for the chemical reaction standard, we had not originally unpacked the standard in terms of atoms and molecules, so we added this interpretation to the unpacking and the learning performances (see Appendix C). Next, we revised the unit to include three learning sets with a focus on substance and property, chemical reactions, and conservation

of mass, with the particle nature of matter integrated throughout each learning set. The unit continuously cycled back to the particle nature of matter with modeling activities, sections of the reader, and discussions after students firsthand experiences of the various phenomena. We also added assessment items targeting the particle nature of matter. This issue demonstrates the critical role of unpacking—the standards themselves often do not clarify what ideas should be brought in as explanations of other ideas, or how deep to take the understanding of a particular idea. In the unpacking process, designers need to develop learning-based arguments for design decisions, which can be revisited if later design cycles so require. As with the previous issue, this feedback encouraged us to reconsider our unpacking, which then influenced the development of the learning performances, curriculum, and assessments. The revised curriculum allowed students to use the new concepts across multiple contexts and phenomena, resulting in a more rigorous treatment of the standards.

Design Issue #4: Students’ Overgeneralizations of Concepts From Exemplars. We found that students sometimes overgeneralized science concepts from exemplars to new contexts where they were not appropriate. One important confusion concerned mixtures and chemical reactions. An example is shown in Table 6, which presents responses to the question “Which change will produce a new substance?” Although burning a candle is the only example of a reaction, the majority of students responded that dissolving lemonade powder in water would produce a new substance. Similarly, a constructed response item described mixing together a clear liquid, white powder and red powder, and asked students to “describe three pieces of evidence you would look for to determine if a chemical reaction occurred.” Here again some students wrote “powder dissolving” would be evidence for a chemical reaction. Students overgeneralized the concept of chemical reactions to include phenomena that were mixtures. We requested a chemistry expert’s feedback on this topic, because the boundaries of what is and is not a chemical reaction are contentious. We wanted to address the students’ conceptions, but also avoid simplifying the complexity of the scientific distinctions in a way that became technically inaccurate.

Considering this potential student misconception made us reconsider how we had unpacked the chemical reaction standard. When we initially unpacked the standard, we focused on what counted as a chemical reaction and did not consider what did not count as a chemical reaction (i.e., boiling and mixing). The data from the first enactment suggested that it was important to consider what is not a chemical reaction in our unpacking. Consequently, we added this idea to the unpacking and made subsequent corresponding changes to the learning performances (see Appendix C) and curriculum materials. We added one lesson to the unit specifically focused on mixtures. Students created a mixture and examined the

TABLE 6
Student Responses to an Item Identifying That a New Substance Formed
(*n* = 89)

Possible Response	Percentage on Pretest	Percentage on Posttest
a. Dissolve lemonade powder in water	55.1	67.4
b. Burning a candle	15.7	16.9
c. Heating water until it evaporates	27.0	14.6
d. Stretching a rubber band	2.2	1.1

properties before and after to determine whether a new substance was made. Furthermore, they analyzed particle models of chemical reactions, phase changes, and mixtures and discussed the similarities and differences of these processes. We also added an explicit section of the student reader to address this issue as well as suggestions for teachers on how to lead discussions around these ideas. Furthermore, this informed how we unpacked content standards in later design. Considering counterexamples in the unpacking of content standards became an important aspect and informed the development of future learning performances, curriculum materials, and assessments.

Design Issue #5: Iterative Alignment of Assessments With Learning Goals. In their initial review of the unit, Project 2061 critiqued a number of the assessment items because they found that the content standards were not necessary and sufficient to complete the items. Later, they performed an extended analysis of a number of our assessment tasks examining five issues: the knowledge needed to answer this task; the target idea being assessed; the knowledge needed to correctly respond; whether the knowledge by itself is enough to correctly respond, or is additional knowledge needed; and whether the task will likely be an effective probe of this knowledge. They found that a number of our assessment items were not aligned with the standards from this particular perspective. We also examined each of the questions on the pre- and posttest to determine whether they aligned with our revised unpacked standards and learning performances. Although when we initially designed the unit, we explicitly tried to align the assessments with the standards, learning performances, and curriculum, the changes we made to these other components made it essential that we go through this process again. Making our learning performances more explicit and looking at the student data from the test, made it apparent that some of the items were not evaluating the intended learning goals.

On the basis of the both our and Project 2061's analysis, we discarded over half of the items on the pre- and posttests. The items that remained were then revised. For example, the question discussed in Design issue #4 was kept because it seemed to address a common student misconception about chemical reactions. But we discarded the choice about stretching a rubber band because so few students selected this choice both before and after the curriculum. We also wrote new items that addressed areas that we were previously missing. For example, we specifically included open-ended question about constructing a scientific explanation for both substances and chemical reactions, so we could examine this scientific practice across different content. This issue illustrates the importance of refining alignment based on multiple cycles of feedback, analysis, and revision.

SECOND ENACTMENT: INFLUENCE OF THE REVISIONS

In this section, we examine student assessment data from the second enactment of the Stuff unit to determine whether our revision using the learning-goals-driven design model promoted greater student learning. We again focus on the chemical reaction learning goals.

Participants and Setting

During the 2002–2003 school year, we scaled up the enactment of the unit to include more teachers, students, schools, and sites than the previous year (Table 7). Nine teachers enacted the unit in three different sites including the two locations from the first enactment (Urban A and Urban B) and one additional location (Large Town C). This enactment included 751 students in seven different schools.

TABLE 7
Teachers, Students, and Classrooms Involved in the Second Enactment

Site	Urban A	Urban B	Large Town C	Total
Schools	4	2	1	7
Teachers	4	2	3	9
Classrooms	14	7	5	26
Students	471	209	71	751

Three of the four schools in Urban A were public middle schools, whereas the fourth school was a charter. Similar to the first enactment, the majority of the Urban A students came from lower to lower-middle income families and over 90% were African Americans. The two schools in Urban B were public middle schools. The students in one school from Urban B came from lower to lower-middle income families. The majority of the students in this school were Hispanic (approximately 82%) with almost half of the students (approximately 46%) identified as English language learners. The second school in Urban B consisted of an ethnically diverse population (approximately 45% Hispanic, 24% Asian, 19% African American, and 12% Caucasian) with students from lower to lower-middle income families. The three teachers in Large Town C were from an independent middle school in a large midwest college town. The majority of these students were Caucasian and from middle to upper-middle income families.

Data Sources

Identical pre- and posttest measures consisted of 30 multiple-choice and 6 open-ended items. Test items measured both the science content standards and scientific inquiry standards addressed in our learning performances. We scored multiple-choice responses for a maximum possible score of 30. We developed specific rubrics to score the six open-ended items with a total maximum score of 30 (McNeill & Krajcik, 2007). All of the open-ended questions were scored by one rater. We then randomly sampled 20% of the tests, and a second independent rater scored them. The average interrater reliability was above 85% for each of the six items. Again, only students who completed both the pre- and posttest assessments were included in the analysis.

Chemical Reaction Achievement by Site

To evaluate the effects of the revisions on the chemical reaction portion of the unit, we examined all the questions on the test that aligned with this content. Table 8 shows the results of the chemical reaction standard for all of the items combined, the items that focused on macroscopic phenomena, and the items that focused on the particle nature of matter. Again, students achieved significant learning gains for this standard. The total effect size for the two urban sites was considerably larger than the first enactment. Large Town C also had a much larger effect size compared to the effect sizes for the two urban sites in the first enactment.

Since we added the particle nature of matter into the chemical reaction component of the unit, we were interested in whether the learning gains were different for the macroscopic phenomena compared to the particle nature of matter. Although there were significant gains for both the macroscopic and the particle model, the effect sizes for the macroscopic

TABLE 8
Enactment 2 Data for Chemical Reactions by Site

Site	Pretest <i>M (SD)</i> ^a	Posttest <i>M (SD)</i>	<i>t</i> -Value ^b	Effect Size ^c
Urban A (<i>n</i> = 244)				
Total chemical reaction	6.84 (2.99)	11.52 (4.33)	18.78***	1.57
Macro	3.34 (2.14)	6.35 (3.04)	15.46***	1.40
Particle	3.50 (1.55)	5.18 (1.84)	14.74***	1.08
Urban B (<i>n</i> = 162)				
Total chemical reaction	5.96 (2.91)	11.36 (4.03)	18.43***	1.86
Macro	2.78 (2.15)	6.61 (2.94)	15.89***	1.78
Particle	3.18 (1.52)	4.75 (1.72)	10.09***	1.03
Large Town C (<i>n</i> = 71)				
Total chemical reaction	9.92 (3.28)	18.04 (2.63)	20.15***	2.48
Macro	5.02 (2.74)	11.39 (2.02)	18.75***	2.32
Particle	4.90 (1.51)	6.65 (1.22)	9.03***	1.16

^aMaximum score: Total = 26, Macro = 17.5, Particle = 8.5.

^bOne-tailed paired *t*-test.

^cEffect size: Calculated by dividing the difference between posttest and pretest mean scores by the pretest standard deviation.

p* < .05; *p* < .01; ****p* < .001.

phenomena were larger across all three sites. Consequently, our next round of revision focused on the particle nature of matter in both the instructional materials and assessment items.

Overall, we observed larger learning gains for the chemical reaction items in the second enactment compared to the first enactment. This suggests that the revision of the materials resulted in greater alignment and support for this content standard.

Revisiting Design Issue #2: Need to Unpack Inquiry Practices

We analyzed whether the changes in the unit resulted in greater student success in constructing scientific explanations (see Table 9).

Overall, students achieved significant gains for all three components of scientific explanation. Again, the effect sizes in the second enactment were larger than the first enactment. Furthermore, the significance of the claim and reasoning learning gains increased compared to the first enactment. Across the three sites, the total explanation score, claim, and reasoning were larger than the effect size the previous year, suggesting that the students had greater success writing scientific explanations with these revised materials.

Revisiting Design Issue #4: Students' Overgeneralizations of Concepts From Exemplars

On the revised pre- and posttests, a similar question was included to assess students' views of mixtures. Again the question gave four scenarios and asked "which will produce new substances?" The frequency of student choices for the pretest was very similar in both enactment #1 and enactment #2 (see Table 10). In the first enactment, more students

TABLE 9
Enactment 2 Data for Scientific Explanations

Site	Pretest <i>M (SD)</i> ^a	Posttest <i>M (SD)</i>	<i>t</i> -Value ^b	Effect Size ^c
Urban A (<i>n</i> = 244)				
Total	1.25 (1.64)	3.13 (2.55)	11.41***	1.15
Claim	0.73 (1.00)	1.42 (1.25)	7.68***	0.69
Evidence	0.42 (0.733)	1.00 (0.98)	8.77***	0.79
Reasoning	0.10 (0.29)	0.71 (0.97)	10.02***	2.10
Urban B (<i>n</i> = 162)				
Total	0.71 (1.39)	3.13 (2.16)	13.84***	1.74
Claim	0.43 (0.86)	1.66 (1.17)	11.23***	1.43
Evidence	0.23 (0.52)	0.67 (0.80)	6.73***	0.85
Reasoning	0.05 (0.27)	0.80 (0.97)	10.19***	2.78
Large Town C (<i>n</i> = 71)				
Total	3.23 (2.52)	6.89 (2.26)	11.42***	1.45
Claim	1.68 (1.28)	2.89 (0.89)	8.10***	0.95
Evidence	1.15 (1.15)	2.08 (1.11)	5.45***	0.81
Reasoning	0.40 (0.71)	1.92 (0.95)	11.68***	2.14

^aMaximum score: Total = 10, Claim = 3.3, Evidence = 3.3, Reasoning = 3.3.

^bOne-tailed paired *t*-test.

^cEffect Size: Calculated by dividing the difference between posttest and pretest mean scores by the pretest standard deviation.

p* < .05; *p* < .01; ****p* < .001.

selected dissolving lemonade after the instructional unit than before, whereas in the second enactment, the number of students selecting dissolving lemonade decreased during the unit. Although the choice decreased, 39% of students on the posttest still selected dissolving lemonade. Adding the lesson on mixtures helped students with this concept, yet it continued to be a difficult area. In the next round of revision, we revisited this section of the instructional materials once again to further address this area of concern.

DISCUSSION

New research-based approaches for instructional materials are needed to support teachers in promoting student learning of the core ideas in science and engage learners in meaningful

TABLE 10
Frequencies for Student Choices on Multiple-Choice Item by Enactment

	Enactment #1 (<i>n</i> = 89) (%)		Enactment #2 (<i>n</i> = 474) (%)	
	Pretest	Posttest	Pretest	Posttest
a. Stretching a rubber band/ hammering metal	2.2	1.1	6.96	1.9
b. Burning a candle	15.7	16.9	22.2	39.0
c. Heating water	27.0	14.6	23.6	19.6
d. Dissolving lemonade	55.1	67.4	47.4	39.0

scientific practices (Duschl et al., 2007; Kesidou & Roseman, 2002). We presented initial evidence that science curriculum materials can embody rigorous treatment of national standards, employ project-based pedagogy in which learners apply content and use scientific practices to solve meaningful problems, and that these materials can result in substantial learning gains.

Developing curriculum materials requires designing materials that respond to a range of interacting concerns, and evaluating these materials relies on multiple aspects of the materials themselves and the factors affecting classroom enactments. Thus, we have utilized approaches from design-based research to identify the elements of our curriculum model that are important in leading to successful outcomes. Although it is impossible to single out the effects of any one aspect of a design process, three aspects of the learning-goals-driven design model have proven useful in identifying design issues and in guiding solutions to these issues: (1) unpacking standards from a learning perspective, (2) learning performances as a way to specify knowledge in use, and (3) the repeated effort to align learning goals with tasks and assessments. In this section, we draw on the specific examples from our iterative development of the Stuff unit to revisit the role of these aspects and discuss how they add to prior design paradigms.

Unpacking Standards From a Learning Perspective

Learning-goals-driven design, like many design approaches, begins with a focus on specifying the goals for learning. Clearly specifying objectives for instruction is a common assumption of instructional design approaches (Dick, Carey, & Carey, 2001; Gagné et al., 2005). The challenge is how to specify these objectives in a way that can productively guide pedagogical designs. While standards have become essential in accountability systems, most curriculum frameworks, selected instructional materials, and assessment systems exhibit alignment only at superficial levels, that is, at the level of topics (Wilson & Berenthal, 2006). We argued that for standards to guide design and assessment, it is essential to unpack these standards from a learning perspective. This unpacking is key to move from an articulation of disciplinary content to learning goals that specify the types of understandings instruction will need to address. We discussed several ways in which unpacking accomplishes the bridge from science content to learning-appropriate goals. The unpacking elaborates the scientific ideas into their constituent aspects and identifies which elements are likely to be problematic for learners, as well as uncovering complexity implicit and not called out in the standard as a separate learning goal. We saw several examples of this in our design iterations. Our analyses identified the concepts of substance and properties as important learning goals, although these ideas were assumed to be transparent in the standards that employed them. We also uncovered the importance of elaborating the chemical reaction standard to rule out combinations that are not chemical reactions. The idea of mixtures emerged as problematic, since students appeared to view any observable change to be evidence of a chemical reaction (e.g., the sugar powder is no longer visible when dissolved in water) instead of focusing on properties of substances changing.

Specifying Learning Goals as Learning Performances

The second key idea in the learning-goals-driven design approach concerns how aspects of the science are specified as learning goals. We build on prior instructional design frameworks by developing a model of objectives based on discipline-specific cognition. We

argued for learning performances as a useful framework, which go beyond general notions of “understanding” to specify the type of cognitive performance that is desired of learners. Thus, rather than “indicators” of learning the concept, these performances are themselves the learning goals. There are many reform initiatives in different countries to embed inquiry in science learning, arguing that these approaches are valuable and yet underrepresented in existing curriculum materials and teaching approaches (Abd-El-Khalick et al., 2004; Black & Atkin, 1996). The use of learning performances enables designers to specify the practices that should be linked to the scientific ideas. Learning performances combine the knowing and doing of science. Thus learning performances provide an approach that is more reflective of the nature of science learning, in which practices and scientific ideas are not separable understandings, but rather interconnected and mutually supportive strands of science literacy (Duschl et al., 2007; Lehrer & Schauble, 2006), rather than the typical approach of specifying separate clusters of content and inquiry goals. Developing learning performances also supports project-based pedagogy, by embedding the commitment to apply science to problems directly into the articulation of learning goals. Furthermore, the learning performances allow us to look at the same content across different inquiry practices and the same inquiry practices across different content to create a more complete picture of a student’s understanding.

Iterative Alignment of Learning Goals, Tasks, and Assessments

A third important aspect of the learning-goals-driven design model is that it provides a process for aligning national standards and learning performances with learning task and assessments. This alignment is consistent with the recent National Research Council report on assessment (Wilson & Berenthal, 2006) that argues for the importance of aligning learning goals with instruction and assessment. The learning-goals-driven design model provides a concrete model that allows curriculum designers and enacting teachers a way to operationalize this recommendation. The learning-goals-driven design model forces designers to use fine-grained alignment in three ways. First, unpacking allows designers to develop a much deeper understanding of what the content and inquiry standards mean and the aspects of understandings entailed in the standard. Second, by developing learning performances, designers must translate science content ideas into cognitive performances, thus carefully specifying the reasoning tasks that require learners to use the content. Hence, learning performances allow designers to create tasks that move beyond a superficial link to the standard to specify the reasoning students should do with that knowledge. Third, unpacking standards and developing performances allows designers to develop tasks that link directly to different types of assessments. Hence, unpacking standards and developing learning performances allow for closer alignment with curriculum tasks and assessments. As we engage in each step, we constantly consider how the different aspects align with each other and the learning goals. Although this emphasis on alignment may seem straightforward, ensuring alignment of theoretical designs with the actual sense teachers and students make of the learning tasks is very challenging, and education reform is rich with examples of design ideas that were implemented quite differently than designers envisioned. Alignment of learning goals and pedagogical approach with materials and assessments is essential for the coherence of the unit as a whole and for student achievement of the learning goals.

In addition to those aspects that arise from the particular nature of the learning-goals-driven design model, we now turn to two issues common to modern design attempts to improve science learning—the use of iterative design and multifaceted feedback. We

examine how these common characteristics of design research have played out in our model.

Feedback and Iterative Design

As many have argued, the challenge of creating quality instructional materials to support teachers and learners in scientific inquiry practices with complex scientific ideas requires iterative design research (Brown, 1992; Edelson, 2002; The Design-Based Research Collective, 2003). Our iterative design efforts draw on multiple data sources. Each data source provides a unique perspective, as well as reinforces the importance of concerns identified from other data sources. For example, the need to unpack inquiry standards resulted from the analysis of pre- and posttests, analyses of student artifacts, and Project 2061's evaluation of the materials.

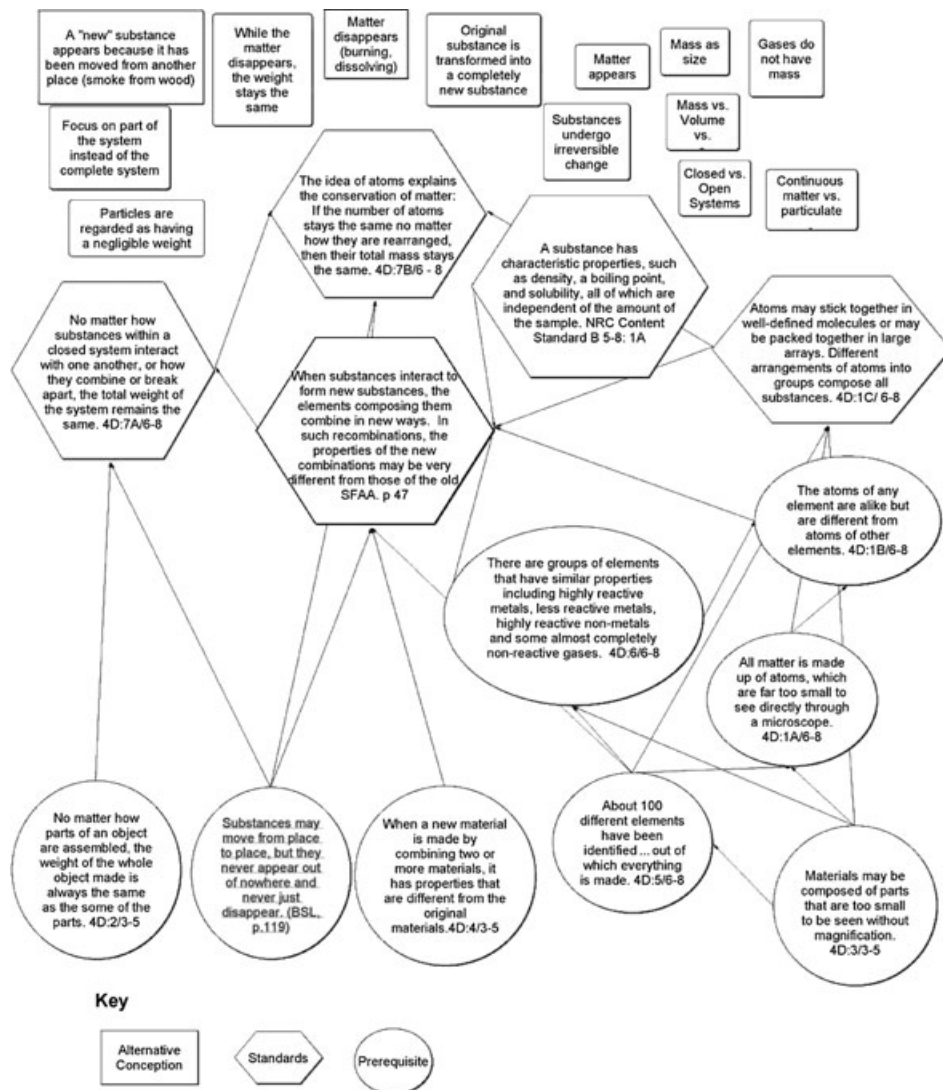
Like other design researchers, our work demonstrates that an ongoing cycle of principled revisions can improve the learning outcome of instructional designs (e.g., Linn & Hsi, 2000). This iterative design process allows designers to blend rigorous treatment of learning goals with the principles of project-based science. Although designers might construct a task or an assessment to help students reach a learning goal, it is only through careful inspection of various data sources and then the reconsideration of the task in light of the learning goal that alignment occurs between learning goals, instruction, and assessment. Knowing what the learning goals mean through unpacking and developing learning performance promotes this process, and is informed through the cycles of design and redesign based on feedback. The explicit unpacking and articulation of learning performances is one way to capture what has been learned from feedback and implemented in the design. Rather than only revising the particular lesson materials, we also capture elaborated understandings of the learning goals in new unpackings of the learning goals, new learning goals added to the unit, new learning performances, and in instructional frameworks like the explanation framework that represents needed support for learners.

These multiple iterations of design allow us to develop tasks that address the driving question of the unit and assessments that align with the learning goal. By tracing each change to the other parts of the materials, we create consistency across the unit including the assessment measures, which are often a neglected portion of the design process. Across the first two enactments of the curriculum materials we found greater learning gains, but we also identified new areas essential for revision. Although we have not discussed the third and fourth rounds of revision and enactment in this paper, we continued engaging in this same process over the next 2 years. Each cycle informs our understanding of the strengths and challenges of the materials and how to make them more successful in real classrooms with teachers and students.

In summary, the approaches of articulating learning goals and pedagogical reforms to bring inquiry into classrooms present core challenges for the field of science education. We suggest that research-based curriculum materials can address these challenges and provide improved tools for learning for teachers and students. To do so, we argue that curriculum materials design should include the three aspects of learning-goals-driven design—unpacking standards from a learning perspective, articulating learning goals as performances that integrate content and practice, and iterative attention to aligning learning goals, tasks, and assessments.

We would like to thank all of the researchers and teachers involved with IQWST and CCMS. We would like specifically to thank Ron Marx and Jo Ellen Roseman for their contributions and feedback on this work.

APPENDIX A: Map of Key Learning Goals for the Stuff Unit



APPENDIX B: Unpacking Standards and Benchmarks

Standard	"Unpacking" the Standard
Science For All Americans (SFAA): When substances interact to form new substances, the elements composing them combine in new ways. In such recombinations, the properties of the new combinations may	Substances have distinct properties and are made of one material throughout. A chemical reaction is a process where new substances are made from old substances. One type of chemical reaction is when two substances are mixed together, and they interact to form new substance(s). The properties of

continued

Standard	"Unpacking" the Standard <i>continued</i>
be very different from those of the old. (AAAS, 1990, p. 47)	the new substance(s) are different from the old substance(s). When scientists talk about "old" substances that interact in the chemical reaction, they call them reactants. When scientists talk about new substances that are produced by the chemical reaction, they call them products.
4D7-Part I No matter how substances within a closed system interact with one another, or how they combine or break apart, the total weight of the system remains the same. (AAAS, 1993)	A closed system is when matter cannot enter or leave a physical boundary. Regardless of how materials interact with each other or change by breaking apart and forming new combinations in a closed system, the total mass of all the material in the system remains the same. The amount of material in our system is represented by the mass of the system. In this case, we are interpreting weight as mass. A common misconception of students is to use mass and weight to have the same meaning. We believe that we need to be consistent in 4D7-Part I and 4D7-Part II. Therefore, we are using the term mass in both Parts I and II.
4D: 1-Part II Atoms may stick together in well-defined molecules or may be packed together in large arrays. Different arrangements of atoms into groups compose all substances. (AAAS, 1993)	Atoms can be arranged in particular ways including the formation of discrete molecules and arrays. A molecule is made up of atoms stuck together in a certain arrangement. An array has repeated patterns of atoms. The different arrangements of atoms give materials different properties. Materials with unique properties are different substances.
4D7-Part II The idea of atoms explains the conservation of matter: If the number of atoms stays the same no matter how they are rearranged, then their total mass stays the same. (AAAS, 1993)	The conservation of matter states that regardless of how substances interact with each other in a closed system, the total mass of all the substances in the system remains the same (4D7-Part I). The majority of substances are made of molecules that are composed of atoms. The reason that the conservation of matter occurs is because the number of atoms of each element in the system stays the same. Regardless of how atoms interact (by breaking apart and reforming new molecules or new arrays) with each other in a closed system, their total mass in the system remains the same.
Added During Design Process	
B 5-8: 1A A substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the sample. (NRC, 1996, p. 154)	Substances have distinct properties that can be used to distinguish and separate one substance from another. Properties such as density, melting point, and solubility, describe the unique characteristics of substances. Density is the mass contained within a unit volume. Melting point is the temperature at which a solid changes to a liquid. Solubility is the ability of a solid to dissolve in a liquid.

APPENDIX C: Revised Learning Performances for Chemical Reactions

Content Standard	Practice (Inquiry Standard)	Learning Performance
Substances react chemically in characteristic ways with other substances to form new substances with different characteristic properties. (NRC, 1996, B:1B/5–8)	<i>Define</i> Provide an accurate scientific description of a concept. (Basic skill—recall)	LP11 Students define a “chemical reaction” as a process in which two or more substances interact [and their atoms combine in new ways] to form new substances with different properties from the old substances. [The new substances are made of the same atoms as the old substances, but the atoms are arranged in new ways.]*
*Molecular Level The idea of atoms and molecules explains chemical reactions: when substances interact to form new substances, the atoms that make up the molecules of the original substances combine in new ways to form the molecules of the new substances. (New AAAS Learning Goal)	<i>Identify</i> Classify an item, phenomenon, or process as an example of a concept. (Basic skill—classify)	LP12 Students identify processes as chemical reactions, phase changes, or mixing.
	<i>Design Investigation</i> Design. . . a scientific investigation. (NRC, 1996, A:1B/5–8) Communicate scientific procedures. . . (NRC, 1996, A:1G/5–8)	LP13 Students design an investigation to determine whether a process is a chemical reaction. They make a prediction, identify variables, control variables, and communicate scientific procedures.
	<i>Conduct Investigation</i> . . . conduct a scientific investigation. (NRC, 1996, A:1B/5–8) Use appropriate tools and techniques to gather. . . data. (NRC, 1996, A:1C/5–8)	LP14 Students conduct a scientific investigation to gather data about properties of substances before and after a process (chemical reaction, phase change, mixing).
	<i>Explain</i> Develop. . . explanations. . . using evidence. (NRC, 1996, A:1D/5–8) Think critically and logically to make the relationships between evidence and explanation. (NRC, 1996, A:1E/5–8)	LP15 Students construct a scientific explanation that includes a claim about whether a process is a chemical reaction, evidence in the form of properties of the substances and/or signs of a reaction, and reasoning that a chemical reaction is a process in which substances interact to form new substances so that there are different substances with different properties before compared to after the reaction.

continued

Content Standard	Practice (Inquiry Standard)	Learning Performance (<i>continued</i>)
	<i>Construct Model</i> Develop. . . models using evidence. (NRC, 1996, A:1D/5–8)	LP16 Students construct molecular models to represent the arrangements of atoms and molecules composing substances before and after a chemical reaction.*
	<i>Use Model</i> Models are often used to think about processes that happen. . . too quickly, or on too small a scale to observe directly. . . (AAAS, 1993, 11B:1A/6–8)	LP17 Students use molecular models of substances before a chemical reaction to reason and represent that during the reaction, two or more substances interact, and their atoms combine in new ways to form new substances. The new substances are made of the same atoms as the old substances, but the atoms are arranged in new ways.* LP18 Students use molecular models of substances before and after a process to identify the process as either: a chemical reaction because the molecular models represent that the atoms composing the old substances before the process combined in new ways to form new substances after the process, or a non-chemical reaction (phase change or mixing) because the molecular models represent the same substances before and after the process.*

*The learning performance targets the content standard at the molecular level.

REFERENCES

- AAAS. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- AAAS. (2001). *Atlas of scientific literacy*. Washington, DC: American Association for the Advancement of Science.
- Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., et al. (2004). Inquiry in science education: International perspectives. *Science Education*, 88(3), 397–419.
- Aikenhead, G. S. (2005). *Science education for everyday life: Evidence-based practice*. New York: Teachers College Press.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Barron, B., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., et al. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences*, 7(3/4), 271–311.

- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22, 797–817.
- Black, P., & Atkin, J. M. (Eds.). (1996). *Changing the subject: Innovations in science, mathematics and technology education*. London: Routledge/OECD.
- Blumenfeld, P., & Krajcik, J. (2006). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 333–354). New York: Cambridge University Press.
- Blumenfeld, P., Soloway, E., Marx, R., Krajcik, J., Guzdial, M., & Palincsar, A. S. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26, 369–398.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Bruozas, M., Finn, L.-E., Tzou, C., Hug, B., Kuhn, L., & Reiser, B. J. (2004). *Struggle in natural environments: What will survive?* In J. Krajcik & B. J. Reiser (Eds.), *IQWST: Investigating and questioning our world through science and technology*. Evanston, IL: Northwestern University.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.
- Dick, W., Carey, L., & Carey, J. O. (2001). *The systematic design of instruction* (5th ed.). New York: Longman.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
- Edelson, D. C. (2001). Learning-for-use: A framework for integrating content and process learning in the design of inquiry activities. *Journal of Research in Science Teaching*, 38(3), 355–385.
- Edelson, D. C. (2002). Design research: What we learn when we engage in design. *The Journal of the Learning Sciences*, 11(1), 105–121.
- Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32(5), 619–634.
- Fensham, P. J., & Harlen, W. (1999). School science and public understanding of science. *International Journal of Science Education*, 21(7), 755–763.
- Fretz, E. B., Wu, H.-K., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Research in Science Education*, 32(4), 567–589.
- Gagné, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design*. Belmont, CA: Wadsworth.
- Harris, C. J., McNeill, K. L., Lizotte, D. J., Marx, R., & Krajcik, J. (2006). Usable assessments for teaching science content and inquiry standards. In M. McMahon, P. Simmons, R. Sommers, D. DeBaets, & F. Crowley (Eds.), *Assessment in science: Practical experiences and education research* (pp. 67–88). Arlington, VA: National Science Teachers Association Press.
- 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(6), 522–549.
- Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research*, 67(2), 227–266.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice. *The Journal of the Learning Sciences*, 12(4), 495–547.
- Krajcik, J., & Czerniak, C. M. (2007). *Teaching science in elementary and middle school: A project-based approach*. Mahwah, NJ: Erlbaum.
- Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy: Supporting development in learning in contexts. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology* (6th ed., Vol. 4). Hoboken, NJ: Wiley.
- Linn, M. C., Bell, P., & Davis, E. A. (2004). *Internet environments for science education*. Mahwah, NJ: Erlbaum.
- Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah, NJ: Erlbaum.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., et al. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, 41(10), 1063–1080.
- McComas, W. F., & Olson, J. K. (1998). The nature of science in international science education standards. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41–52). Dordrecht, The Netherlands: Kluwer.

- McLaughlin, M. W., & Shepard, L. A. (1995). Improving education through standards-based reform. A report by the National Academy of Education panel on standards-based education reform. Stanford, CA: National Academy of Education, Stanford University.
- McNeill, K. L., Harris, C. J., Heitzman, M., Lizotte, D. J., Sutherland, L. M., & Krajcik, J. (2004). How can I make new stuff from old stuff? In J. Krajcik & B. J. Reiser (Eds.), *IQWST: Investigating and questioning our world through science and technology*. Ann Arbor, MI: University of Michigan.
- McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 233–265). Mahwah, NJ: Erlbaum.
- McNeill, K. L., & Krajcik, J. (in press). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2), 153–191.
- Millar, R., & Osborne, J. (1998). *Beyond 2000: Science education for the future*. London: King's College.
- Moje, E. B., Peek-Brown, D., Sutherland, L. M., Marx, R., Blumenfeld, P., & Krajcik, J. (2004). Explaining explanations: Developing scientific literacy in middle-school project-based science reforms. In D. Strickland & D. E. Alvermann (Eds.), *Bridging the gap: Improving literacy learning for preadolescent and adolescent learners in grades 4–12* (pp. 227–251). New York: Teachers College Press.
- NRC. (1996). *National Science Education Standards*. Washington, DC: National Research Council.
- OECD. (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris: Organisation for Economic Cooperation and Development.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020.
- Perkins, D. (1998). What is understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice* (pp. 39–58). San Francisco: Jossey-Bass.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, 13(3), 337–386.
- 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: Erlbaum.
- Rivet, A., & Krajcik, J. S. (2004). Achieving standards in urban systemic reform: An example of a sixth grade project-based science curriculum. *Journal of Research in Science Teaching*, 41(7), 669–692.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5–51.
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3), 345–372.
- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28, 859–882.
- Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, 37(5), 525–559.
- 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(3), 283–312.
- Sherin, B. L., Edelson, D., & Brown, M. (2004). On the content of task-structured curricula. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education* (pp. 221–248). Dordrecht, The Netherlands: Kluwer.
- Singer, J., Marx, R. W., Krajcik, J., & Chambers, J. C. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist*, 35, 165–178.
- Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233–267). Washington, DC: Falmer Press.
- Spillane, J. P. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge, MA: Harvard University Press.
- Tal, T., & Kedmi, Y. (2006). Teaching socioscientific issues: Classroom culture and students' performances. *Cultural Studies of Science Education*, 1(4), 615–644.
- The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8.
- Tomorrow 98. (1992). Report of the superior committee on science, mathematics and technology education in Israel. Jerusalem: The Ministry of Education and Culture.

- Tyler, R. W., Gagné, R. M., & Scriven, M. (1967). *Perspectives of curriculum evaluation*. Chicago: Rand McNally.
- Wiggins, G. P., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wilson, M. R., & Berenthal, M. W. (2006). *Systems for state science assessment*. Washington, DC: National Academies Press.
- Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1), 35–62.