*American Educational Research Journal*
*Month XXXX, Vol. XX, No. X, pp. 1–34*
*DOI: 10.3102/0002831215585962*


# An Efficacy Trial of Research-Based Curriculum Materials With Curriculum-Based Professional Development

Joseph A. Taylor
*Abt Associates Inc.*
Stephen R. Getty
*Colorado College*
Susan M. Kowalski
Christopher D. Wilson
*Biological Sciences Curriculum Study*
Janet Carlson
*Stanford University*
Pamela Van Scotter
*Biological Sciences Curriculum Study*

*This study examined the efficacy of a curriculum-based intervention for high school science students. Specifically, the intervention was two years of research-based, multidisciplinary curriculum materials for science supported by comprehensive professional development for teachers that focused on those materials. A modest positive effect was detected when comparing outcomes from this intervention to those of business-as-usual materials and professional development. However, this effect was typical for interventions at this grade span that are tested using a state achievement test. Tests of mediation suggest a large treatment effect on teachers and in turn a strong effect of teacher practice on student achievement—reinforcing the hypothesized key role of teacher practice. Tests of moderation indicate no significant treatment by demographic interactions.*

KEYWORDS: curriculum materials, curriculum-based professional development, efficacy trial, high school science, inquiry, BSCS 5E Instructional Model

## The Current State of Curriculum Efficacy Research

Science education is at a critical juncture where evidence of the effects of curriculum materials is greatly needed. It is not enough to develop curriculum materials and professional development (PD) programs whose

component features are based on extant research and hope they are effective. Many stakeholders in education, practitioners in particular, need evidence from rigorous trials about which comprehensive programs (i.e., year-long programs using multiple, integrated features) have the greatest effects on student outcomes. Although numerous federal agencies have funded the development of curriculum materials over the past 30 years, the field of science education still lacks evidence regarding what programs (or types of programs) have noteworthy effects. As a consequence, school and district decision makers have had little guidance toward implementing potentially more efficacious programs that might displace those that have smaller or no effects (Slavin, 2008) and no way to appropriately respond to the need for improved STEM education.

The Institute for Education Sciences (IES) established the What Works Clearinghouse (WWC) to provide stakeholders in education with information on programs that have undergone rigorous efficacy trials. The WWC provides stakeholders such as school districts information with which to make evidence-based decisions related to instructional interventions. At

---

Joseph A. Taylor is a principal associate/scientist at Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138; e-mail: *Joseph_Taylor@abtassoc.com*. He was director of research and development at Biological Sciences Curriculum Study (BSCS) when this research was conducted. His research focuses on studying the impact of curriculum materials and professional development on student outcomes and the policy implications of such studies.

Stephen R. Getty is the director of the Quantitative Reasoning Center at Colorado College in Colorado Springs, Colorado. His research in education focuses on methods to measure student motivation in STEM disciplines and faculty/teacher interventions to increase student motivation.

Susan M. Kowalski is a senior science educator at BSCS. Her research focuses on developing and studying science curriculum and professional development programs in online contexts.

Christopher D. Wilson is senior science educator at BSCS. His research focuses on the measurement of teacher and student learning in science education, measuring teacher pedagogical content knowledge with human and automated scoring, and the impact of lesson analysis–based professional development.

Janet Carlson is an associate professor (research) and the director of the Center to Support Excellence in Teacher (CSET) in the Graduate School of Education at Stanford University. She was the executive director at BSCS when this research was conducted. Her research focuses on the interplay between transformative professional development, the enactment of curriculum, teacher practice, and student learning.

Pamela Van Scotter is currently senior associate director at BSCS, where she oversees the programmatic work in curriculum development, professional development, and research and evaluation. She served as principal investigator for the project that developed the instructional materials that were the subject of this study. She has a background in anthropology and linguistics.

the time this article was written, just eight studies of science interventions had met the evidence standards of the WWC (IES, 2014), and only one is related to curriculum materials or PD for use at the high school level.

This lack of evidence is particularly troubling as we find ourselves at a time of rapid reform. As the new *Framework for K–12 Science Education* (National Research Council [NRC], 2012) and the *Next Generation Science Standards* (*NGSS*; NGSS Lead States, 2013) have been released, teachers across the country will be expected to teach new disciplinary core ideas, crosscutting concepts, and scientific practices. While national documents clearly describe the three dimensions of *NGSS* well, minimal empirical evidence accompanies these documents about the role of research-based curriculum materials in supporting student attainment of standards as well as the nature of the PD programs most likely to help teachers implement effective instruction. It is therefore imperative that curriculum and PD programs that attempt to, or claim to, support such goals are subjected to rigorous trials that can make confident causal claims about their impacts.

## Study Overview

This study sought to test the causal link between a curriculum-based science education intervention and increased student achievement. The primary goals of the research were to (a) test the overall efficacy of research-based curriculum materials with associated PD for improving high school science achievement, (b) explore the role of teacher practice in the relationship between use of the curriculum materials and improved student achievement outcomes, and (c) explore the extent to which treatment effects were equitable across demographic groups. The curriculum materials under study are titled *BSCS Science: An Inquiry Approach* (hereafter referred to as *An Inquiry Approach*). The materials were developed with funding from the National Science Foundation (ESI 9911614 and ESI 0242596)

## Theoretical Underpinnings of Research-Based Curriculum Materials

In this article, we define curriculum materials to include both the student text as well as teacher support materials. This section addresses the theory behind both the student and teacher curriculum materials.

### The Role of Curriculum Materials in Science Classrooms

Curriculum materials can be a means to improve student interest and achievement in science (NRC, 2007). The notion that curriculum materials truly matter and directly influence the learning process has been supported in the literature for decades (e.g., Forbes & Davis, 2010; Schmidt, McKnight,

& Raizen, 1997; Usiskin, 1985). Curriculum materials play a defining role in classrooms, affecting both what and how teachers teach (NRC, 2002). Ball and Cohen (1996) explain this powerful influence:

> Unlike frameworks, objectives, assessments, and other mechanisms that seek to guide curriculum, instructional materials are concrete and daily. They are the stuff of lessons and units, of what teachers and students do. . . . Not only are curriculum materials well-positioned to influence individual teachers' work but, unlike many other innovations, textbooks are already "scaled up" and part of the routine of schools. They have "reach" in the system. (p. 6)

Further, Schmidt, Houang, and Cogan (2002) caution against efforts to improve instruction that are isolated from efforts to improve curriculum materials available to teachers and students. "If we pretend that the textbook doesn't exist—and conduct PD in ways that assume teachers will implement an entirely different approach to content than the texts take—believe me, the textbook will win" (p. 18).

### Constructivism and Student Learning

Constructivism is a key foundation that frames research-based curriculum materials designed to emphasize opportunities for students to develop conceptual understandings of science. Our work is based on two common theoretical bases for constructivist research: Ausubelian theory (Ausubel, Novak, & Hanesian, 1978) and the work of L. S. Vygotsky (1978). Ausubelian theory states that a learner's prior knowledge is an important factor in determining what is learned in a given situation. Vygotsky's work emphasizes the relationship between the teacher's prior knowledge and the students' prior knowledge as well as the importance of the social construction of knowledge. Students and teachers may use similar words to describe concepts yet have very different personal interpretations of those concepts. Vygotsky's work implies that science curriculum and instruction should take into account the differences between teacher and student conceptions and should provide time for student-to-student interaction so that learners can develop concepts from those whose understandings and interpretations are closer to their own.

Discussions of constructivist teaching and learning have been hampered by inconsistency in how it is envisioned by different scholars and researchers, including being equated with completely unguided or "discovery" learning. Hmelo-Silver, Duncan, and Chinn (2007) discuss the range of perspectives and offer this definition of a constructivist learning environment: an environment in which "students are cognitively engaged in sensemaking, developing evidence-based explanations, and communicating their ideas. The teacher plays a key role in facilitating the learning process and may provide content knowledge on a just-in-time basis" (p. 100). It is this interplay

between the student as sensemaker and the teacher as facilitator that defines our view of constructivist learning environments.

Today we see the roots of constructivism reflected in comprehensive reviews of the literature on learning, such as *How People Learn* (NRC, 2000). The authors summarize three key ideas about learning, suggesting that students come to the classroom with preconceptions that shape their learning, student competence requires a deep foundation of knowledge as well as an understanding of how this knowledge relates to a framework, and students benefit from explicitly monitoring and taking control of their own learning. The *Inquiry Approach* curriculum materials incorporated into the study intervention were strongly influenced by these findings.

### Coherence, Focus, and Rigor

Though more than a decade old, the findings from the Trends in Mathematics and Science Study (TIMSS) analysis (Schmidt et al., 2001) and the research synthesis *How People Learn* (NRC, 2000) provide clear and compelling guidance for the development of effective curriculum materials. In general, these reports indicate that curriculum materials in the United States need to be more focused by having a storyline organized around key concepts, more coherent by having explicit connections between ideas, and more rigorous by setting high standards for learners with respect to both their cognitive and metacognitive development. Curriculum materials developed within a framework that is coherent both within years and across years facilitates a deeper student conceptual understanding (American Association for the Advancement of Science [AAAS], 2001; Carlson, Davis, & Buxton, 2014; NRC, 1999, 2007). Yet, persistent evidence indicates that curriculum materials in science are fragmented, lacking coherence, and not well articulated through a sequence of grade levels (AAAS, 2001; Kesidou & Roseman, 2002; Schmidt et al., 1997, 2001; Schmidt, Wang, & McNight, 2005). As a result, curriculum materials in the United States generally cover many concepts, often repeating concepts annually without depth (Schmidt et al., 1997, 2001). Most materials focus on details that are tangential to the key ideas and fail to make connections across units when the same key idea is presented in different ways (Kesidou & Roseman, 2002).

### Educative Materials and Teacher Support in the Classroom

Research-based curriculum materials for students will never eliminate the important role of the teacher in the classroom. Teachers ultimately shape how curriculum materials are enacted in the classroom (Beyer & Davis, 2012; Forbes & Davis, 2010). Teachers select elements of text to include for instruction, and they emphasize or deemphasize aspects of a curriculum based on their own understanding and beliefs about what is best for students. Remillard (2005) described the complex teacher-curriculum relationship as

contextually based, dependent on both the teacher and the curriculum, and tightly interconnected with other teaching practices. If a teacher's understandings and beliefs about instruction align with the philosophy of the curriculum, then it is likely that there will be a synergistic relationship between use of the materials and practice (Powell & Anderson, 2002). On the other hand, a teacher may understand instruction and hold beliefs about practices that diverge from the philosophy of the materials, creating a gap between what curriculum developers intended and what the teacher actually enacts in the classroom (Ball & Cohen, 1996).

Because of the ubiquitous placement of curriculum materials in the school setting, there is unique potential for curriculum materials to support teachers as learners. Teachers may use their curriculum materials to deepen their content knowledge, gain ideas for how to present complex information to students, or determine how they might assess student learning. Some researchers describe curriculum materials that explicitly address the teacher as learner as "educative" (Beyer, Delgado, Davis, & Krajcik, 2009; Davis & Krajcik, 2005). Davis and Krajcik (2005) identified nine heuristics to describe educative science materials and how science curriculum materials can support teachers' enactment of reform-based instruction with their students. The heuristics focus on teacher subject matter knowledge as well as pedagogical content knowledge (PCK). In addition, they articulate the importance of including a rationale for curricular design decisions and providing supports for teachers to adapt materials. Specifically, embedded educative teacher resources can include information and rationale on the instructional model, additional scientific background, alternative understandings students may have associated with the content, as well as suggestions for enhancing students' abilities to function as a group. As such, educative science curricula become a resource for teachers, supporting them as they use materials in their own instructional settings.

Schneider and Krajcik (2002) observed that teachers who use educative materials enhance their content learning and better implement specific strategies and representations suggested in the materials. Similarly, Davis and Krajcik (2005) noted a link between teachers' use of educative curriculum materials for science and their PCK for corresponding science topics.

### Theoretical Underpinnings of Curriculum-Based Professional Development

Although educative curriculum materials have clear advantages, they are often complex, and teachers benefit from additional support to fully understand the curricula they are trying to implement (Davis & Krajcik, 2005). For example, teachers often do not fully avail themselves of the information provided in the teacher's materials. In addition, even when teachers do study the educative materials, they will likely interpret the information using their own experiential lenses (McNeill, 2009). Thus, the evidence suggests that integrating educative materials with face-to-face PD could be the most

effective approach to enhancing teachers' understanding of the philosophy and key features of curriculum materials. This should be particularly true when the PD incorporates elements of effective PD practice. Loucks-Horsley, Hewson, Love, and Stiles (2003) reviewed extensive research on the characteristics of effective PD and identified several that are particularly germane to PD aimed at supporting the use of curriculum materials. These effective practices include providing coherent, ongoing (i.e., multi-event) programs that mirror best practice; keeping a focus directly on learning and teaching; and providing teachers opportunities to develop deep understanding of concepts and participate in communities of reflective practice. Last, when teachers adopt research-based curriculum materials, it is essential that they learn about key features of the materials as well as the rationale for why developers incorporated those key features (Lin & Fishman, 2006). These features include the instructional model for the materials (Penuel, Gallagher, & Moorthy, 2011) as well as key elements such as building an inquiry culture, using sensemaking strategies, and understanding the storylines for each unit of instruction. The concept of educative materials for teachers supported by complementary PD merges two areas of research: (a) the role of educative materials and (b) the numerous reports indicating that PD focused on the implementation of well-designed materials can have a significant impact on teaching and learning (Briars & Resnick, 2001; Darling-Hammond, 1997; Heller, Daehler, Shinohara, & Kaskowitz, 2004; Ladewski, 1994; Powell & Anderson, 2002; Schneider & Krajcik, 2002).

## The Resulting Experience for Students Using *An Inquiry Approach*: The Treatment Condition

### Curriculum Materials

These three major theoretical underpinnings for curriculum materials (constructivism, coherence, and educativeness) formed the foundation for the design of *An Inquiry Approach*. As a result, this high school program aims to support not only the learning of science concepts but also the development of a culture of learning that empowers students and teachers to learn science and conduct scientific inquiry.

#### *Constructivism*

One of the primary ways that the materials attend to the research on constructivism is by structuring learning around the BSCS 5E Instructional Model (Bybee, 1997; Bybee & Landes, 1990). In particular, the BSCS 5E (engage, explore, explain, elaborate, and evaluate) Instructional Model supports the teacher in scaffolding the learning experiences for students and provides a research-based, social constructivist storyline throughout each chapter. The BSCS 5E Instructional Model organizes the instructional

*Table 1*
**How Curriculum Materials Embody the Principles of Learning**

| Principles of Learning From *Knowing What Students Know*[a] | Related Features of the Materials |
|---|---|
| Instruction is organized around meaningful problems and goals. | Activities center on relevant problems and/or current research with an inquiry focus. |
| Instruction must provide scaffolds for solving meaningful problems and supporting learning for understanding. | Instruction follows the 5E Instructional Model; scaffolding is particularly strong in the explore/explain cycles. |
| Instruction must provide opportunities for practice with feedback, revision, and reflection. | The materials include both formative and summative assessments as well as metacognitive strategies aligned with the activities. |
| The social arrangements of instruction must promote collaboration and distributed expertise as well as independent learning. | The materials include an appropriate mix of small team activities, partner projects, jigsaws, presentations, class discussions, and individual work. |

[a]Pellegrino, Chudowsky, and Glaser (2001).

sequence so that students have multiple opportunities to develop a deep understanding of concepts through practice, feedback, revision, and reflection. See Table 1 for details.

In the initial lesson of each chapter, there are opportunities for the students to consider, express, represent, and share their current understanding about a concept. This is a critical opportunity for students that sets the stage for and promotes learning (NRC, 2000). This in turn helps the teacher frame the subsequent lessons and lab activities where students have the opportunity to explore questions in small teams. These experiences result in a set of common experiences on which teams will continue to build their understanding.

Exploratory lab activities lead to other lessons and interactive readings that help students formulate and articulate their foundational understanding. Activities designed to reinforce and expand students' understanding follow. These lab activities often ask the students to test their understanding in a different setting or by adding a new variable. Many of the lessons also provide opportunities for collaborative learning that model the scientific enterprise. Such work in heterogeneous groups promotes the back-and-forth process essential to knowledge construction (Vygotsky, 1962).

The role of formative assessment in these curriculum materials is important for both the students and the teachers (Atkin, 2002; Black & Wiliam, 1998; NRC, 2001). During each lesson, the students complete tasks and respond to questions that serve as benchmarks for students and teachers to assess their learning

experiences. In addition, students have multiple opportunities to develop explanations from evidence using appropriate scaffolds (McNeill & Krajcik, 2007). At the end of a chapter, students complete a comprehensive lab activity or other type of lesson designed to demonstrate their understanding for themselves and for their teacher with respect to core concepts presented in the chapter. This experience serves as a summative assessment for the chapter.

The instructional model in these materials supports both the teacher and the student in creating a culture of inquiry in the classroom and provides opportunities for students to develop an understanding of science by practicing science (NGSS Lead States, 2013) and reflecting on its nature, their experiences, and their findings. Through a coherent sequence of these lab activities and questions aimed at improving critical thinking skills, the students, over time, have opportunities to learn explicitly about the nature of science in the context of learning rigorous content.

### *Coherence*

The feature of coherence in curriculum materials is critical to student learning (Rutherford, 2000; Schmidt et al., 1997) and foundational to the curriculum materials used during the intervention. Because we know that students are more likely to learn best when their learning experiences are grounded within a coherent conceptual framework (NRC, 2000), the development of such a framework was a critical first step in the design of the curriculum materials used in this study. The use of a framework facilitated the development of a focused, conceptual storyline within each chapter and across chapters, resulting in a coherent learning experience for students. Each of the core units comprises four chapters, with each unit exposing students to fundamental concepts in one of the science disciplines (i.e., physical science, life science, earth science, science and society), and this multidisciplinary cycle repeats in Grades 10 and 11 (see framework in Table 2). The last chapter in each core unit allows students to apply what they have learned thus far in an integrated context. The result of this articulation among units is that the number of major concepts students learn is fewer and the experiences with those concepts are more sophisticated, with the goal that students' understanding of them is deeper and more complex (Schmidt et al., 2001).

To further promote coherence in *An Inquiry Approach*, the development team used the *Understanding by Design* process (Wiggins & McTighe, 2005). This process is divided into three main stages. In Stage I, Desired Results, the team identified the enduring understandings for students based on the national standards (the *National Science Education Standards* at that time) and scientific expertise. In Stage II, Assessment Evidence, the team developed the assessment tasks that would serve as evidence that the students had gained the targeted understandings. In Stage III, Learning Plan, the team developed the sequence of learning experiences that were hypothesized to

*Table 2*

**Curricular Framework for *An Inquiry Approach* With *Next Generation Science Standards* Alignment**

| Units | Major Concepts Addressed at Each Grade Level | | |
|---|---|---|---|
| | 9 | 10 | 11 |
| Science as Inquiry (also integrated throughout each content area) | • Questions and concepts that guide scientific investigations | • Design of scientific investigations<br>• Communicating scientific results | • Evidence as the basis for explanations and models<br>• Alternative explanations and models |
| Physical Science | • Structure and properties of matter<br>• Structure of atoms<br>• Integrating chapter (DCI PS1.A, PS1.C) | • Motions and forces<br>• Chemical reactions<br>• Integrating chapter (DCI PS1.B, PS2.A, PS2.C, PS3.A, PS3.B) | • Interactions of energy and matter<br>• Conservation of energy and increase in disorder (DCI PS3.A) |
| Life Science | • Cell structure and function<br>• Behavior of organisms<br>• Integrating chapter (DCI LS1.A, LS1.B) | • Biological evolution<br>• Molecular basis of heredity<br>• Integrating chapter (DCI PS3.D, LS1.C-D, LS3.A-B, LS4.A-C, LS4.E, ETS1.A) | • Matter, energy, and organization in living systems<br>• Interdependence of organisms<br>• Integrating chapter (DCI LS1.C) |
| Earth-Space Science | • Origin and evolution of the universe<br>• Origin and evolution of the Earth system<br>• Integrating chapter (DCI PS1.B, PS4.A-C, ESS1.A-C, ESS4.B) | • Geochemical cycles<br>• Integrating chapter (DCI ESS2.A-B, ESS3.D) | • Energy in the Earth system (DCI PS3.C, ETS1.B) |
| Science in a Personal and Social Perspective; Science & Technology | • Personal and community health<br>• Natural and human-induced hazards<br>• Abilities of technological design (DCI PS4.C, LS2.A-B, LS2.C) | • Population growth<br>• Natural resources<br>• Environmental quality (DCI LS4.D, ESS3.C, ESS4.A, ESS4.C, ETS 2.B) | • Science and technology in local, national, and global challenges<br>• Understandings about science and technology (DCI ESS3.C, ETS2.A) |

*Note.* Codes beneath main concepts show the alignment of the program with the Disciplinary Core Ideas (DCI) in the *Next Generation Science Standards* for physical sciences (PS), life sciences (LS), earth and space sciences (ESS), and engineering, technology, and applications of science (ETS).

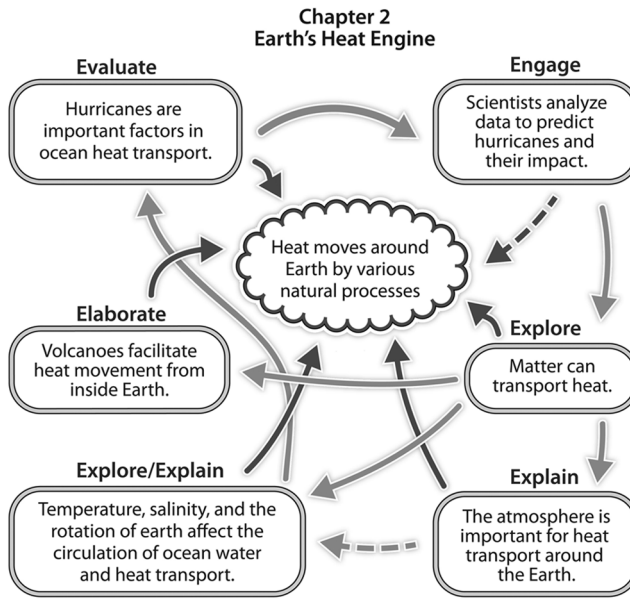


*Figure 1.* **Example of a conceptual flow graphic (CFG). The dark arrows represent connections to the central concept of the chapter. The lighter arrows represent connections among ideas through the sequence of activities. Dashed arrows indicate weaker connections.**

enable students to construct the targeted understandings and be successful on the assessments. Also, in the third stage, developers made extensive use of conceptual flow graphics (CFGs). These CFGs are visual diagrams that illustrate the flow of ideas through the chapter and the relative strength of conceptual connections among them (see Figure 1). Use of CFGs throughout the development process refined and strengthened the focus of the materials, in turn strengthening both rigor and coherence.

These processes for developing the materials positioned the program well with the NGSS, helping to ensure coherence within a unit and across grade bands. For example, the *Inquiry Approach* program supports engaging students in the practices of science in many ways. The BSCS 5E Instructional Model that organizes and sequences instruction provides opportunities for students to ask questions about phenomena, design simple experiments, use evidence collected from their classroom experiences to develop explanations, and communicate scientific ideas to their peers. Thus, engaging students in the practices of science is integral to the structure of the program. Further, the concepts in Table 2 clearly align with the

Disciplinary Core Ideas (DCI) of the Next Generation Science Standards (NGSS) for high school science. In Table 2, codes beneath main concepts show the alignment of the program with the DCI in the NGSS for physical sciences (PS), life sciences (LS), earth and space sciences (ESS), and engineering, technology, and applications of science (ETS). Finally, Table 3 illustrates an example of how one crosscutting concept is woven across multiple units and years of the program. All seven crosscutting concepts are woven throughout all three years of the program.

To make the materials educative, we integrated a variety of teacher supports that align with the heuristics of educative materials suggested by Davis and Krajcik (2005). Included in the support materials of *An Inquiry Approach* are background on the philosophy behind the instructional model; practical strategies for implementing the instructional model as intended; strategies to support meaningful, collaborative learning; and strategies for empowering students to monitor and support their own learning (e.g., through the effective use of student notebooks). The teacher materials also include additional content background for teachers who find themselves teaching outside their area of expertise. This includes information on common conceptions students have about specific concepts and how to best address them (see positive relationship between teacher knowledge in this area and student outcomes in Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). It also includes specific ideas for both formative and summative assessment of student learning.

In sum, *An Inquiry Approach* is a comprehensive set of curriculum materials intended to promote the types of learning just described. By *comprehensive*, we mean that each level of the program supports teachers and students for a full year of high school science content without the need for supplementation and is designed to be used every day for the entire school year. *An Inquiry Approach* is a three-year program, typically used in Grades 9 through 11. However, because the desired outcome measure was administered in the spring of students' 10th-grade year, the efficacy study was limited to estimating the program's effects after just two years of use. The treatment condition also included a seven-day PD program for teachers, provided each year of the curriculum program to support implementation. The PD program is described in the following section.

## The Professional Development Program

In the context of the curriculum-based PD provided in this study, we translated recommended practices into a PD program that engaged teachers in a year-long experience with a clear focus on student learning and the effective implementation of the program. The providers of the PD program were also developers of the curriculum materials. In addition, they had

*Table 3*

**Examples of How the *Inquiry Approach* Program Addresses One of the Seven *Next Generation Science Standards* (*NGSS*) Crosscutting Scientific Concepts**

| *NGSS* Crosscutting Concept | Disciplinary Units | Grade 9 | Grade 10 |
|---|---|---|---|
| Systems and system models[a] "Defining the system under study—specifying its boundaries and making explicit a model of that system—provides tools for understanding and testing ideas that are applicable throughout science and engineering." (NRC, 2012) | Physical Science | Students consider the movement of matter into and out of systems during chemical reactions. | Students consider the transfer of energy within a system during collisions in both microscopic and macroscopic models. |
| | Life Science | Students examine cells as systems with important matter and energy inputs and outputs and cells as components of larger structures in plant and body. systems. | Students simulate population systems using models of predator-prey interactions. |
| | Earth-Space Science | Students simulate the transfer of electromagnetic radiation into and out of a system using springs and ropes to model starlight with different frequencies and amplitudes. | Students use physical models and chemical reactions to understand that systems on Earth have reservoirs of matter, inputs and outputs, and fluxes of matter between those reservoirs. |
| | Science and Society | In several investigations, students analyze inputs and outputs of and processes within ecosystems and how fire suppression policies affect forest ecosystems. | Students model how human activities modify natural processes and alter the distribution of species and resources in ecosystems. |

[a]The other six crosscutting concepts are (a) patterns; (b) cause and effect; (c) scale, proportion, and quantity; (d) energy and matter; (e) structure and function; and (f) stability and change.

extensive experience providing PD on this curriculum program. The research team also attended each PD session and monitored the PD implementation, providing input as necessary.

The goals of PD were to maximize implementation fidelity by helping teachers deepen their understanding of the nature of the materials by modeling lessons, encouraging collaboration around common experiences with the materials, improving teachers' content knowledge, as well as enhancing their ability to implement the instructional model that organizes and sequences all instruction in *An Inquiry Approach*. The seven-day PD program each year was composed of a three-day summer institute and four one-day sessions throughout the school year. The extended duration of the PD enabled us to work with the teachers throughout the year, particularly at the beginning of each new unit, which focused on a different science discipline. The extended duration also allowed us to introduce new features of the program as teachers' understanding of the program expanded and their comfort level with previously introduced features increased. Thus, in this study, the face-to-face PD sessions complemented the educative aspects of the teacher support materials and aimed to provide teachers with the experiences needed to take full advantage of research-based curriculum materials. Teachers in the PD program were engaged as collaborative learners of content with the PD facilitators from BSCS. The PD providers used (or approximated) the pedagogical methods suggested in the program for students. As teachers engaged in activities as science learners, the activities became the common experience that anchored subsequent conversations of pedagogy.

## The Comparison Condition

During the two years of the study, the comparison group continued to use their own extant curriculum materials and received the usual PD planned by their schools and districts (i.e., business as usual [BaU]). In Grade 9, teachers in the comparison schools used one of eight different textbooks provided by their school districts. Most of the eight textbooks were each used by only one or two comparison teachers. However, the *Prentice Hall Physical Science* and *Earth Science* textbooks were used by over half of the 28 comparison teachers (16 teachers in one of the larger participating school districts). In Grade 10, most BaU schools progressed to their standard 10th-grade biology curriculum and textbooks. That said, a one-unit sample of artifacts collected by external researchers indicated that BaU teachers used their textbooks only 24% of the time, indicating that they supplemented the district-supplied textbook with many other curriculum materials.

*Table 4*
**Demographics by Treatment Group**

| Demographics by Group | Treatment ($n$ = 1,509) | Comparison ($n$ = 1,543) |
|---|---|---|
| Female | 747 | 722 |
| Special education | 166 | 143 |
| Free/reduced lunch | 796 | 592 |
| English language learners | 94 | 70 |
| American Indian | 38 | 18 |
| Asian | 82 | 169 |
| Black | 121 | 87 |
| Hispanic | 454 | 379 |
| Hawaiian/Pacific Islander | 8 | 15 |
| Mixed race/ethnicity | 19 | 81 |
| White | 787 | 794 |
| Number of suburban schools | 4 | 5 |
| Number of rural schools | 5 | 4 |

## Theory of Change

Our theory of change is that the combination of educative curriculum materials for teachers, research-based materials for students, and curriculum-based PD will produce a positive effect on both students and teachers and that the effect on students is in part mediated by positive effects on teachers' practice. More specifically, research-based student materials provide scaffolding for exemplary teacher practice while the educative teacher materials and face-to-face PD provide the necessary supports for teachers to enact that curriculum within their own contexts.

## Study Description

### Setting and Participants

The study reported here took place in 18 high schools (9 treatment, 9 comparison) in the state of Washington. These 18 schools initially enrolled in the study for the 2009–2010 school year. Approximately half of the 18 schools were in suburban areas. The remaining schools were in rural areas. Table 4 provides an overview of student and school characteristics for the treatment and comparison groups. Both groups were somewhat diverse in terms of student demographics, and each group had a similar blend of suburban and rural schools. Differences in student demographics across groups were accounted for in the analyses.

### Study Eligibility

All traditional high schools that had not used *An Inquiry Approach* in the past and that participated in the state achievement testing program were eligible for the study. Further, for a school to be eligible, the principal agreed to encourage teacher attendance at PD sessions and to allow access to classrooms for data collection.

### Research Questions

This study was guided by a primary research question related to the efficacy of *An Inquiry Approach* plus curriculum-based PD as well as a set of exploratory questions related to mediation and moderation of treatment effects.

*Research Question 1:* Primary analyses of treatment effects on student achievement: Controlling for covariates, what is the main effect of treatment on student achievement?

*Research Question 2:* Exploratory analyses of mediation: To what extent does teacher practice mediate the effect of treatment on student achievement?

*Research Question 3:* Exploratory analyses of moderation (interactions): To what extent do student demographic characteristics moderate the effect of treatment on students (i.e., what are the interaction effects of treatment with student characteristics)?

### Design

This study uses a pretest/posttest control group design (Shadish, Cook, & Campbell, 2002) where the means of posttreatment outcome measures are compared across the treatment and comparison groups after being controlled for pretreatment differences in outcomes. The unit of random assignment to groups was the school (or "cluster" of students), and as such, the design is also often referred to as a *cluster-randomized trial* (Raudenbush, 1997). Neither matching nor blocking was used prior to random assignment as the late timing of schools joining the study made it impossible for reliable stratification levels to be established.

### Group Allocation and Attrition

As two of the developers of *An Inquiry Approach* were also on the research team, a number of safeguards were employed to limit experimenter bias. The first such safeguard was using an external researcher to make random assignments to groups using a random number generator. The researcher did not make the assignments until schools had consented to participate in the study, and no schools left the study after learning of their group assignment. We used information in Table 5 to determine how

equivalent the groups were in science achievement (pretreatment). Specifically, we compared treatment and comparison group means on the 8th-grade science baseline covariate for the baseline and analytic samples, respectively. The baseline sample was the set of students in schools who were randomly assigned to groups at the onset of the study. The analytic sample was the set of students in schools randomly assigned at the onset of the study for which a posttreatment (Grade 10) science outcome measure was available. By comparing the baseline and analytic sample sizes within Table 5, the degree of outcome measure attrition is apparent. The overall attrition rate of individual students in the study, based on availability of the 10th-grade science outcome score, is 18%, with an attrition rate of 18% in the treatment group and 17% in the comparison group. Thus, the differential attrition rate across groups is 1%. Consulting Table 5, it is clear that the random assignment process was not completely successful in distributing baseline achievement levels evenly across groups. The baseline difference in the 8th-grade state science achievement scores across groups was noteworthy (Hedges' $g$ = .23) and in favor of the comparison group. This baseline difference was accounted for in the treatment effect models described later in this article. Finally, we note here that the primary analysis for the main effect of treatment is an *intent-to-treat* analysis. Thus, all students retained their original treatment group assignment in the analytic sample (i.e., students crossing groups during the intervention were treated in the analysis as if they remained in their original treatment group for the full two years).

## Measures

In this section, we describe the two measures used in the analysis. The first is the outcome measure used to estimate the main effect of treatment on student achievement. The second is a measure of classroom practice and culture that was used in the mediation analysis.

### *The Outcome Measure*

In this study, we sought an outcome measure with four key features: (a) importance to all stakeholders, including teachers, students, parents, and administrators; (b) alignment with the student abilities and understandings that the *Inquiry Approach* curriculum materials seek to improve; (c) a fair outcome measure for the comparison group; and (d) strong psychometric properties. For this study, the outcome measure chosen was the Washington state science assessment (the High School Proficiency Exam; HSPE), which clearly meets the first criterion. As for the second criterion, the HSPE has reasonable alignment with the *Inquiry Approach* program due to its broad coverage of science content (earth/space, physical, and life sciences) as well as its focus on science practices (e.g., developing questions and designing investigations, evidence as the basis for explanations and models, and communicating scientific results). We

*Table 5*

**Pretreatment Sample Sizes and Characteristics for the Baseline and Analytic Samples in a Two-Level Cluster-Randomized Trial**

| | Treatment Group | | | Comparison Group | | |
|---|---|---|---|---|---|---|
| | Sample Size | Sample Characteristics | | Sample Size | Sample Characteristics | |
| | Students/ Schools | Individual/School Mean | Standard Deviation | Students/ Schools | Individual/School Mean | Standard Deviation |
| **Baseline sample** | | | | | | |
| Individual baseline measure | | | | | | |
| Eighth-grade science achievement test | 1,845 | 385.96 | 27.816 | 1,868 | 391.85 | 27.544 |
| School baseline measure | | | | | | |
| Eighth-grade science achievement test | 9 | 386.15 | 9.27 | 9 | 388.55 | 8.73 |
| **Analytic sample** | | | | | | |
| Individual baseline measure | | | | | | |
| Eighth-grade science achievement test | 1,509 | 388.84 | 27.379 | 1,543 | 394.92 | 26.549 |
| School baseline measure | | | | | | |
| Eighth-grade science achievement test | 9 | 388.21 | 8.76 | 9 | 390.58 | 10.21 |

concluded that the third criterion would be met by the HSPE as this test is based on Washington state standards to which all schools (treatment and comparison) are held accountable. The psychometric properties of the HSPE are strong. For the 10th-grade science outcome measure, the reported Cronbach's alpha is .87. For the baseline achievement covariates—the state achievement test scores for 7th-grade writing, 8th-grade math, and 8th-grade science—the alpha values are .77, .90, and .89, respectively (Education Testing Service, 2012).

### *The Teacher Practice Measure*

We used the Reformed Teaching Observation Protocol (RTOP; Piburn et al., 2000; Sawada et al., 2002) as the primary measure of teacher practice. From this point forward, all references to "teacher practice" should be read as "teacher practice as indicated by the RTOP." The RTOP instrument measures the extent to which science and mathematics teaching aligns with the recommendations for research-based instructional reform described in national science and mathematics standards documents of the late 1990s. The instrument is made up of 25 Likert-type items, divided into five subscales: (a) Lesson Design & Implementation, (b) Content—Propositional Knowledge, (c) Content—Procedural Knowledge, (d) Classroom Culture—Communicative Interactions, and (e) Classroom Culture—Student-Teacher Relationships. A total score across all items is also calculated. Each scale varies from a score of 0, *behavior never occurred*, to a score of 4, *pervasive or extremely descriptive of the lesson*. As a whole, the protocol addresses teacher attention to students' prior knowledge, student engagement in a learning community, and the extent to which teachers support an atmosphere of problem solving and student-generated ideas. Validation studies of the RTOP suggest that it can have strong psychometric properties. The reliability estimate ($R^2$) for the entire instrument is .954 (Piburn et al., 2000).

To guard against experimenter bias, two external researchers conducted the classroom observations. To calibrate the observers before the site visits, the two observers watched classroom instruction from video recordings. After each video, the observers then independently provided ratings across the 25 RTOP items. After this "pre-discussion" rating, the two observers discussed their ratings and the basis for their ratings.

Classroom observations were made throughout the year. Nearly all teachers in this study were observed eight times (approximately once each month). We chose this comprehensive approach to increase the likelihood that the average RTOP score for each teacher was representative of his or her typical practice. In the analyses, the outcome measure for teachers is their mean RTOP score across their observations. The external observers of teacher practice were never told of teachers' treatment group assignments. However, it likely became discernible as the students used their *Inquiry Approach* textbooks that are designed to be used most days, if not every

day, in class. Therefore, we acknowledge that we cannot rule out the possibility of observer bias in the ratings of teacher practice (RTOP scores).

Because two external researchers each scored one half of the treatment and comparison classrooms, interrater reliability was calculated to test for consistency in scoring between raters. A sample of 7.4% of the observations was scored by both raters (29 out of a total of 394), and interrater reliability was calculated using the intraclass correlation coefficient (ICC). Analysis of the commonly scored observations yielded an intraclass correlation coefficient of .966 (total RTOP scores, two-way mixed effects model, absolute agreement, average measures). Interpretation of the intraclass correlation coefficient is similar to that of Cohen's kappa, where a commonly used rule of thumb is that .40 to .59 represents moderate interrater reliability, .60 to .79 is substantial, and greater than .80 is outstanding (Landis & Koch, 1977).

The same external researchers also used a Fidelity of Implementation Observation Protocol (BSCS, 2009) to assess treatment teachers' use of the instructional materials with students. In particular, the tool examined the quality of teachers' use of the BSCS 5E Instructional Model. The average score on this protocol across 183 independent observations of treatment teachers was 2.13 on a 3-point scale (71%), indicating overall program use consistent with the developers' intent.

## Analysis and Findings

### Confirmatory Analyses: Main Effect of Treatment on Students

The primary purpose of the study was to address Research Question 1: Controlling for covariates, what is the main effect of treatment on student achievement? Because assignment to treatment or comparison (BaU) conditions occurred at the school level while the outcome of interest occurs at the student level, we chose multilevel modeling to estimate the effects of the treatment on student achievement. Preliminary analyses confirmed this modeling choice as the effect of clustering is sizeable (unconditional ICC = .13), and the data meet the multilevel modeling assumptions of homogeneity of Level 1 variance ($\chi^2$ = 22.87, $p$ = .153) and normality of residuals (Q-Q plots of residuals at both levels have patterns that are generally linear).

To refine the estimate of the treatment effect, we included in the model a set of covariates that we hypothesized to be correlated with the outcome measure: students' scores on the 10th-grade state science assessment (SCI10). This set of covariates included both demographic and achievement variables. Some student-level (Level 1) covariates were also used in aggregate at the school level (Level 2). The Level 1 covariates included achievement scores such as those from the 8th-grade state science assessment (SCI8), the 8th-grade state math assessment (MAT8), and the 7th-grade state writing assessment (WRIT7). The Level 1 demographic covariates included free and

reduced-price lunch status (FRL) as a proxy for socioeconomic status, gender (GEND), English language learner status (ELL), special education status (SPED), grade level (GRADE), and a set of race contrast codes that include American Indian (AMIND), Asian (ASIAN), Black (BLK), Hispanic/Latino (HISP), Hawaii/Pacific Islander (HPI), and those who indicated two or more ethnicities (MIX), each of which allows achievement comparisons between the selected group of students and the reference group of White students.

The Level 2 model included the treatment variable (TREAT) as well as a school mean aggregate for the eighth-grade science assessment score (MNSCI8), a school mean aggregate for the eighth-grade math assessment score (MNMAT8), and a school mean aggregate for the FRL status (MNFRL). Because we had large numbers of Level 1 units (students) in the sample but a relatively small number of Level 2 units (just 18 schools), we were much more judicious in including Level 2 covariates, including only the most theoretically influential on the outcome, as each of these consumes a degree of freedom in the Level 2 statistical significance tests. We grand-mean centered all independent variables to facilitate the desired covariate adjustment. The main effect model was specified and run as described next using STATA 12 statistical software.

Level 1:

$$\begin{aligned} SCI10_{ij} = {} & \pi_{0j} + \pi_{1j}(\mathrm{SCI8})_{ij} + \pi_{2j}(\mathrm{MAT8})_{ij} + \pi_{3j}(\mathrm{WRIT7})_{ij} + \pi_{4j}(\mathrm{FRL})_{ij} \\ & + \pi_{5j}(\mathrm{GEND})_{ij} + \pi_{6j}(\mathrm{ELL})_{ij} + \pi_{7j}(\mathrm{SPED})_{ij} + \pi_{8j}(\mathrm{GRADE})_{ij} \\ & + \pi_{9j}(\mathrm{AMIND})_{ij} + \pi_{10j}(\mathrm{ASIAN})_{ij} + \pi_{11j}(\mathrm{BLK})_{ij} + \pi_{12j}(\mathrm{HISP})_{ij} \\ & + \pi_{13jk}(\mathrm{HPI})_{ij} + \pi_{14j}(\mathrm{MIX})_{ij} + e_{ij}. \end{aligned}$$

Level 2:

$$\begin{aligned} \pi_{0j} = {} & \beta_{00} + \beta_{01}(TREAT)_j + \beta_{02}(MNSCI8)_j \\ & + \beta_{03}(MNMAT8)_j + \beta_{04}(MNFRL)_j + r_{0j}. \end{aligned}$$

Descriptive statistics for the outcome variable (SCI10) are provided in Table 6. The main effect estimates from the multilevel model are provided in Table 7.

This output suggests that the treatment group students would have scored an estimated 3.68 scale score points higher, on average, than students in the comparison group had the groups been fully equivalent prior to treatment. This difference ($\beta_{01}$) is statistically significant at the $\alpha = .05$ significance level ($p = .035$).

Like many efficacy trials, this study was subject to attrition and a resulting loss of data. As a result, the research team replicated this treatment effect analysis after imputing the missing data using a multiple imputation algorithm within STATA 12. Specifically, within STATA, we used the EM Algorithm with multiple imputations to address missing data. The treatment

*Table 6*
**Posttreatment Outcomes for the Analytic Sample and Estimated Effects in a Two-Level Cluster-Randomized Trial**

| | Treatment Group | | Comparison Group | | Estimated Effects | |
|---|---|---|---|---|---|---|
| Outcome Measure | Covariate-Adjusted Mean | Unadjusted Standard Deviation | Covariate-Adjusted Mean | Unadjusted Standard Deviation | Covariate-Adjusted Mean Difference | *p* Value |
| 10th grade science | 384.52 | 43.305 | 380.84 | 39.119 | 3.68 | .035 |

*Table 7*
**Estimates of Fixed Effects on 10th-Grade Science Achievement Score**

| Independent Variable | Level | Coefficient | Standard Error | *z* Value | *p* Value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| SCI8 | Student | 0.71 | 0.03 | 26.14 | <.001 | 0.65 | 0.76 |
| MAT8 | Student | 0.32 | 0.02 | 16.13 | <.001 | 0.28 | 0.35 |
| WRIT7 | Student | 0.80 | 0.32 | 2.50 | .012 | 0.17 | 1.43 |
| FRL | Student | –2.61 | 1.01 | –2.59 | .010 | –4.58 | –0.63 |
| GEND | Student | –5.41 | 0.90 | –6.02 | <.001 | –7.17 | –3.65 |
| ELL | Student | –8.86 | 2.20 | –4.02 | <.001 | –13.17 | –4.54 |
| SPED | Student | –4.70 | 1.62 | –2.90 | .004 | –7.88 | –1.52 |
| GRADE | Student | –11.96 | 2.54 | –4.72 | <.001 | –16.93 | –6.99 |
| RACE-AMIND | Student | –6.92 | 3.28 | –2.11 | .035 | –13.36 | –0.48 |
| RACE-ASIAN | Student | –1.70 | 1.70 | –1.00 | .317 | –5.05 | 1.64 |
| RACE-BLK | Student | –5.31 | 1.82 | –2.92 | .004 | –8.87 | –1.74 |
| RACE-HISP | Student | –5.38 | 1.29 | –4.18 | <.001 | –7.90 | –2.86 |
| RACE-HPI | Student | –10.25 | 5.36 | –1.91 | .056 | –20.76 | 0.26 |
| RACE-MIX | Student | –3.92 | 2.47 | –1.59 | .113 | –8.75 | 0.92 |
| TREAT | School | 3.68 | 1.75 | 2.11 | .035 | 0.25 | 7.10 |
| MNSCI8 | School | 0.34 | 0.34 | 1.00 | .318 | –0.32 | 1.00 |
| MNMAT8 | School | 0.08 | 0.23 | 0.34 | .734 | –0.38 | 0.54 |
| MNFRL | School | –2.89 | 7.89 | –0.37 | .715 | –18.34 | 12.57 |

effect estimate from the identical model applied to the imputed data sets yielded very similar results ($\beta_{01}$ = 3.37, *SE* = 1.65, *p* = .041), suggesting that the missing data did not introduce a systematic bias in the treatment effect estimate.

As an additional way to interpret the treatment effect, the research team also computed the effect size. The Hedge's *g* effect size (with small sample size adjustment $\omega$), a measure of practical significance, was computed by using the treatment effect coefficient ($\beta_{01}$) from the multilevel model as

the covariate-adjusted mean difference across groups (the numerator). The denominator was the pooled standard deviation weighted for sample size differences across groups:

$$g = \frac{\omega\beta_{01}}{\sqrt{\frac{(n_i-1)s_i^2+(n_c-1)s_c^2}{n_i+n_c-2}}}.$$

The Hedge's *g* value for the treatment effect was .09 standard deviations. The 95% confidence interval for the effect size is [.01, .17]. The initial power analysis was conducted using Optimal Design Plus Empirical Evidence (v.3) software (Raudenbush et al., 2011). In this analysis, we assumed an unconditional ICC of .15, a Level 2 covariate correlation ($R^2$ value) of .50, 18 schools, and 150 students per school. These values corresponded to a minimum detectable effect size of $d$ = .40. In actuality, the power was significantly better than we anticipated, allowing us to detect a much smaller effect than expected. First, the actual ICC was lower than our initial estimate (unconditional ICC = .13), and the average number of students per school was higher ($n$ = 170). In addition, the analysis included covariates at both Level 1 and Level 2 of the model, and the Level 2 covariates provided us with much better precision than we anticipated ($R^2_{L2}$=0.91). As a result, we had sufficient power in our study to detect a smaller effect size than we initially expected.

The WWC would characterize an effect size of .09 as a "statistically significant positive effect." Although the WWC reserves the characterization of "substantively important" for effects larger than .25, this effect is meaningful in the context of the small effect sizes often observed in high school interventions. For example, Hill, Bloom, Black, and Lipsey (2007) conducted a synthesis of effect sizes for randomized control trials and found that for elementary school studies, the average effect size was .33, and for high school studies, the average effect size was .27. However, this high school average included effect sizes computed on outcome measures that were proximal or targeted to the intervention. Based on what Hill and colleagues observed in the elementary school studies, inclusion of proximal effect sizes likely inflated the average effect size for high school interventions. Specifically, they found that the average effect size varied by the breadth of focus for the outcome measure, reporting that "within studies of elementary schools, mean effect sizes are highest for specialized tests (0.44), next-highest for narrowly focused standardized tests (0.23), and lowest for broadly focused standardized tests (0.07)" (p. 8). Given that the effect size reported for this study (.09) was computed using scores from a broadly focused standardized test (HSPE 10 Science) and is associated with a high school intervention, we find the effect size to be within expectation.

Another way to interpret this effect size is to compare it to normative expectations for achievement growth (i.e., average pre-post year effect sizes for 9th- and 10th-grade science students). This effect size expresses students' expected gain in science achievement over the course of one year. Looking across a set of nationally normed tests, Bloom and colleagues (2008) estimated that the average pre-post year effect size for science in 9th grade is .19 and .22 for 10th grade. Thus, the two-year expected gain in achievement can be estimated as .41 standard deviations. The effect size of .09 detected in this two-year intervention study is noteworthy as it corresponds to .09/.41 or 22% of the two-year expected gain. Multiplying .22 by 18 school months for a two-year intervention, we estimate that treatment group students emerge from the study (i.e., start 11th grade) nearly four months ahead of comparison group students in science achievement.

As a final way to express the practical importance of the treatment effect, we converted the effect size into an improvement index using the properties of the normal distribution. In a normal distribution, a 1.0 *SD* effect size is equivalent to 34 percentile points. Therefore, an effect size of .09 equates to an improvement index of 3.06 (34 × .09) percentile points. So, if the comparison students were at the mean of the normed sample, the 50th percentile, the treatment group students would then be placed at just over the 53rd percentile.

There are two key reasons why these estimates of the true treatment effect are likely conservative. First, we join other researchers who have noted that using new interventions that require unfamiliar practices can often lead to an "implementation dip," where use of the program features is mechanistic and can result in a *negative* effect on outcomes for some time prior to ultimately improving outcomes (e.g., Fullan, 2001; Hall & Hord, 2001). *An Inquiry Approach* encourages teachers to use instructional practices that are not commonplace in high schools (Banilower et al., 2013), and as such, these practices were likely unfamiliar to a majority of treatment teachers. This implementation dip, if a factor in this study, would reduce the size of the treatment effect. In contrast, many comparison teachers were using programs or sets of activities that they used routinely prior to the research. Second, observations indicate that the learning experiences of students in the comparison group included some research-based practices similar to those promoted in the treatment group. This would also tend to reduce the treatment effect.

Other main effects in Table 7 are interesting as well. The achievement covariates were all highly predictive of the SCI10 outcome measure. In general, the race covariates have statistically significant ($\alpha = .05$) main effects on achievement with the exception of the race dummy codes that compare the achievement of Asian, Hawaiian/Pacific Islander, and mixed ethnicity students to White students, respectively. In this sample, males had higher mean scores than females, economically advantaged students had higher

mean scores than economically disadvantaged students, native language English speakers had higher mean scores than English language learners, and students without a special education designation had higher mean scores than those with such a designation. In a later section, we report whether the relationships between key demographic variables and achievement described here differ by treatment group.

### Exploratory Analyses: Mediation and Indirect Treatment Effects

The purpose of the mediation analysis was to address Research Question 2: To what extent does teacher practice mediate the effect of treatment on student achievement? The research team hypothesized that the nature of teachers' practice is critical to the efficacy of the curriculum materials and that improving teacher practice is part of the mechanism by which the causal effect of the treatment is realized. This is supported by syntheses of intervention studies such as that conducted by Nye, Konstantopoulos, and Hedges (2004), who observed that the proportion of variance in student outcomes attributable solely to *between teacher* variance can be as much as 20%. As a result, this efficacy study sought to use the RTOP to collect comprehensive data about teacher practices to test whether the treatment (curriculum materials plus PD) has an *indirect* effect on students' science achievement via teacher practice as a mediating variable (see Figure 2).

In Figure 2, path *a* represents the effect of the treatment on teacher practice (RTOP); path *b* represents the effect of teacher practice (the mediator) on the science achievement outcome (SCI10), controlling for the treatment; and path *c′* represents the effect of the treatment on the science achievement outcome (SCI10), controlling for teacher practice. That is, *c′* is the direct (unmediated) effect of treatment. The product of paths *a* and *b* is often used to represent the mediating (indirect effect) of the treatment on the outcome (MacKinnon, 2008).

The mediation design for this study is often referred to as a $3 \longrightarrow 2 \longrightarrow 1$ design because the treatment is at the third level (school), the mediator is measured at the second level (teacher), and the outcome is measured at the first level (student). We tested mediation in this $3 \longrightarrow 2 \longrightarrow 1$ design using a modeling approach advocated by leading methodologists (MacKinnon, 2008; Pituch, Murphy, & Tate, 2010). In this approach, separate equations for the mediator and outcome can be used to estimate the indirect effect. The first set of equations in the following estimates path *a*. The teacher-level equation for the mediator is

$$RTOP_{ij} = \pi_{0j} + r_{0ij},$$

where $RTOP_{ij}$ is the teacher-level mediator, $\pi_{0j}$ is the RTOP mean for school $j$, and $r_{0ij}$ is the teacher-level random effect. The school-level equation is

$$\pi_{0j}=\beta_{00}+\beta_{01}TREAT_j+u_{0j},$$

where $\beta_{01}$ is the effect of the treatment on the RTOP scores (path $a$ of Figure 2), and $u_{0j}$ is the school-level random effect. The main effect of treatment on teacher practice estimated from these models is $\beta_{01}$ = 16.74 ($SE$ = 3.11, $p <$ .001), corresponding to raw group means and standard deviations of 71.4 (10.1) and 55.0 (7.6) for treatment and BaU, respectively. This is a large effect with corresponding Hedge's $g$ value = 1.85. Estimating the $b$ and $c'$ paths requires a three-level model. The student-level equation for the outcome is

$$SCI10_{ijk}=\pi_{0jk}+e_{ijk},$$

where $\pi_{0jk}$ represents the outcome mean for teacher $j$ of school $k$, and $e_{ijk}$ is the student-level random effect. The teacher-level equation adds the mediator (RTOP) as a predictor:

$$\pi_{0jk}=\beta_{00k}+\beta_{01k}(RTOP)_{jk}+r_{0jk},$$

where $\beta_{01k}$ represents the within-school impact of RTOP on the mean SCI10 score. The school-level equations are

$$\beta_{00k}=\gamma_{000}+\gamma_{001}(TREAT)_k+u_{00k} \text{ and } \beta_{01k}=\gamma_{010},$$

where $\gamma_{001}$ is path $c'$ of Figure 2 and $\gamma_{010}$ is the fixed effect of RTOP on SCI10 (controlling for treatment), or path $b$. The fixed effects from this three-level model are $c'$ ($\gamma_{001}$) = 1.56 ($SE$ = 2.21, $p$ = .49) and $b$ ($\gamma_{010}$) = 0.13 ($SE$ = 0.07, $p$ = .07). The presence of a strong treatment effect on the teacher-practice mediator, a nearly significant association between teacher practice and student achievement and a small, remaining direct treatment effect ($c'$), is consistent with our mediational hypothesis. A more formal test is described in the following.

The indirect effect of teacher practice can be estimated as the product of the $a$ and $b$ paths or the $ab$ product. This product is ($\gamma_{01}$)($\gamma_{010}$) = (16.74)(0.13) = 2.18. The 95% confidence interval for the indirect effect was computed using the PRODCLIN Program (MacKinnon, Fairchild, & Fritz, 2007), yielding [–0.12, 4.84] and a corresponding probability of type 1 error ($p$ = .064). Although mediation is sometimes considered to be present

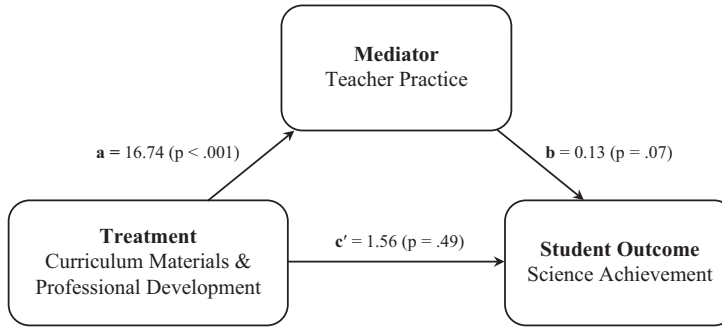


*Figure 2.* **Mediation of the treatment effect with unstandardized coefficients.**

only when the *ab* product is statistically significant at $\alpha = .05$, we consider this result suggestive of mediation and worthy of discussion.

Direct and indirect effects can be parsed using the following expression:

$$\tau = \tau' + ab,$$

where $\tau$ is the total effect (3.68) from the main effect of treatment analysis, $\tau'$ is the direct (unmediated) effect (1.56) from the mediation analysis, and *ab* is the indirect effect of teacher practice on student achievement (2.18).

From this decomposition of effects, we can estimate that the indirect or mediation effect of teacher practice is (2.18/3.68) or 59% of the total effect of the treatment. The magnitude of this proportion supports our hypothesis that teacher practice truly matters in the implementation of the curriculum program. Note that for this mediation test and for the moderation tests that follow, there was no adjustment to the significance level for multiple hypothesis tests as these tests are framed as exploratory.

### Exploratory Analyses: Testing Moderation Through Interaction Effects

The purpose of the moderation analysis was to address Research Question 3: To what extent do student demographic characteristics moderate the effect of treatment on students (i.e., what are the interaction effects of treatment with student characteristics)? Given that nearly all the demographic variables had significant main effects on achievement, it became important to take a comprehensive approach to testing whether these main effects were maintained when students were disaggregated by treatment group. Thus, we specified a random slopes model to test the cross-level (treatment at Level 2, demographics at Level 1) interactions between

treatment and all of the demographic variables of interest. The results of the cross-level interaction analyses were mixed (i.e., some positive, some negative), but none of the effects were statistically significant at the $\alpha = .05$ level.

## Discussion and Implications

Efficacy studies such as the one described in this article are urgently needed (Hmelo-Silver et al., 2007). For example, there is an ongoing debate about whether students are better served by direct instruction or constructivist approaches to learning (Kirschner, Sweller, & Clark, 2006; Tobias & Duffy, 2009). Klahr (2010) asserts "the burden of proof is on constructivists to define a set of instructional goals, an unambiguous description of instructional processes, a clear way to ensure implementation fidelity, and then to perform a rigorous assessment of effects" (p. 4). Some constructivists have expressed resistance to direct rigorous comparisons of these different instructional approaches, arguing that due to fundamental differences between constructivist pedagogies and direct instruction, no common research method can evaluate the two (Jonassen, 2009). Alternatively, Klahr states, "Constructivists cannot use complexity of treatments or assessments as an excuse to avoid rigorous evaluations of the effectiveness of an instructional process" (p. 3). Similarly, Mayer (2004) recommends that we "move educational reform efforts from the fuzzy and unproductive world of ideology—which sometimes hides under the various banners of constructivism—to the sharp and productive world of theory-based research on how people learn" (p. 18).

Toward these challenges, this study adds to a small, extant set of rigorous studies on the effects of curriculum interventions based in research on constructivist learning (e.g., Clements & Sarama, 2008; Lynch, Pyke, & Grafton, 2012). In this study, we conclude that the combination of research-based curriculum materials and curriculum-based PD was effective, observing a positive treatment effect on students' science achievement. This finding is consistent with other recent studies of interventions that combine curriculum materials and PD (e.g., August et al., 2014; Domitrovich et al., 2009). The size of the effect in this study is within an expected range for high school interventions and for when a broadly focused outcome measure (state achievement test) is used.

This efficacy study is unique in its formal test of the effect of curriculum materials that simultaneously embody several theoretical frameworks, including constructivism; educative curriculum materials; and curriculum coherence, focus, and rigor. Further, as a constructivist learning model was integral to *An Inquiry Approach*, findings from this study meet the request of scholars such as Klahr and Mayer by serving as examples of rigorous evidence in support of constructivist approaches. In the same vein, the findings of this study challenge the calls for more direct instruction made by

Kirschner et al. (2006), Stull and Mayer (2007), and Kirschner and van Merriënboer (2013).

The primary treatment effect detected in this study is necessarily limited in specificity as the multifaceted nature of this intervention and the nature of the study design prevented us from isolating the unique effects of discrete program features. For example, in this study it was not possible to disentangle the unique effects of formative assessment from the effects of the instructional model or the effects of educative materials from those of face-to-face PD. We think that testing the efficacy of individual features of this intervention is useful for some features but not for others as some are too interrelated with other features to be adequately isolated.

Further, the results of tests for whether the intervention led to more equitable outcomes or "outputs" (Lynch, 2001) for students remain inconclusive. Some treatment-demographic interaction effects suggested more equitable outcomes for the treatment group while others suggested more equity in the BaU comparison group. None of the interaction effects were statistically significant, so it is not entirely clear whether the equity-focused features of the materials had systematic effects on achieving equitable outcomes across demographic groups.

On the other hand, more compelling results come from the exploratory mediation analysis where we observed a strong treatment effect on teacher practice (RTOP) and a positive teacher practice effect on student achievement. This mediation result suggests that (a) teaching practice can be improved by educative, research-based instructional materials in concert with face-to-face, curriculum-based PD and (b) teaching practice indeed matters over and above the inherent features of the curriculum materials for students.

Mediation results such as these could have larger implications, especially if the role of teacher practice continues to be observed as highly influential in future efficacy studies of curriculum programs. Specifically, if it is widely observed that the effects of research-based curriculum interventions tend to be nonsignificant once the influence of teacher practice is controlled, what does that mean for *effectiveness* and *scale-up* studies as defined by the *Common Guidelines for Education Research and Development* (IES/NSF, 2013)? In this document, each of these types of impact studies includes a test of effects under routine (non-idealized) conditions. If routine conditions were to mean the removal of PD support and the results of this study prove to be consistent with that of the field at large, it seems unlikely that many interventions that require significant teacher expertise to implement will produce positive effects. We suggest then that the field consider a notion of curriculum-based interventions where corresponding PD is a standard program feature and not an upgrade that can be disregarded or deemed as optional to successful implementation. Acknowledging that adding PD support will increase the cost to school districts of implementing research-

based programs, we suggest from the results of this study that the additional cost is a worthwhile investment.

## Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant number R305K060142 to BSCS. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We also thank the Office of Assessment and Student Information at the OSPI in Washington for providing student demographic and achievement data. Finally, the authors would like to recognize Karen Askinas for her important contributions to this research.

## References

American Association for the Advancement of Science. (2001). *Atlas of science literacy*. Washington, DC: AAAS and National Science Teachers Association.

Atkin, J. M. (2002). Using assessment to help students learn. In R. W. Bybee (Ed.), *Learning science and the science of learning book* (pp. 97–103). Arlington, VA: NSTA Press.

August, D., Branum-Martin, L., Cardenas-Hagan, E., Francis, D., Powell, J., Moore, S., & Haynes, E. (2014). Helping ELLs meet the Common Core State Standards for Literacy in Science: The impact of an instructional intervention focused on academic language. *Journal of Research on Educational Effectiveness*, *7*(1), 54–82.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York, NY: Holt, Rinehart and Winston.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–14.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education*. Chapel Hill, NC: Horizon Research, Inc.

Beyer, C. J., & Davis, E. A. (2012). Learning to critique and adapt science curriculum materials: Examining the development of preservice elementary teachers' pedagogical content knowledge. *Science Education*, *96*(1), 130–157.

Beyer, C. J., Delgado, C., Davis, E. A., & Krajcik, J. (2009). Investigating teacher learning supports in high school biology curricular programs to inform the design of educative curriculum materials. *Journal of Research in Science Teaching*, *46*(9), 977–998.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–148.

Bloom, H., Zhu, P., Jacob, R., Raudenbush, S., Martinez, A., & Lin, F. (2008). *Empirical issues in the design of group-randomized studies to measure the effects of interventions for children* ( MDRC Working Papers on Research Methodology). Retrieved from http://www.mdrc.org/publication/empirical-issues-design-group-randomized-studies-measure-effects-interventions-children

Briars, D., & Resnick, L. (2001). *Standards, assessments—and what else? The essential elements of standards-based school improvement*. Paper presented at the Local Systemic Change PI Meeting, National Science Foundation, Washington, DC.

BSCS. (2009). *Fidelity of implementation protocol*. Retrieved from www.bscs.org.

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

Bybee, R., & Landes, N. (1990). Science for life and living: An elementary school science program from Biological Sciences Curriculum Study. *The American Biology Teacher*, *52*, 92–98.

Carlson, J., Davis, E. A., & Buxton, C. (2014). *Supporting the implementation of the* Next Generation Science Standards (NGSS) *through research: Curriculum materials*. Retrieved from https://narst.org/ngsspapers/curriculum.cfm

Clements, D., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443–494.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. New York, NY: National Commission on Teaching and America's Future.

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

Domitrovich, C., Gest, S., Gill, S., Bierman, K., Welsh, J., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, *46*(2), 567–597.

Education Testing Service. (2012) *Washington Comprehensive Assessment Program Grades 3–8, High School Spring 2011 technical report*. Retrieved from http://www.k12.wa.us/assessment/pubdocs/WCAP2011SpringAdministrationTechnicalReport.pdf

Forbes, C. T., & Davis, E. A. (2010). Curriculum design for inquiry: Preservice elementary teachers' mobilization and adaptation of science curriculum materials. *Journal of Research in Science Teaching*, *47*(7), 820–839.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.

Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Boston, MA: Allyn and Bacon.

Heller, J. I., Daehler, K. R., Shinohara, M., & Kaskowitz, S. R. (2004). *Fostering pedagogical content knowledge about electric circuits through case-based professional development*. Paper presented at the National Association for Research on Science Teaching (NARST) Annual Meeting, Vancouver, WA.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). *Empirical benchmarks for interpreting effect sizes in research*. Retrieved from http://www.mdrc.org/publicaitons/459/full.pdf

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107.

Institute for Education Sciences. (2014). *What Works Clearinghouse*. Retrieved from http://www.ies.ed.gov/ncee/wwc/Topic.aspx?sid=14

Institute for Education Sciences and The National Science Foundation. (2013). *Common Guidelines for Education Research and Development*. Retrieved from http://ies.ed.gov/pdf/CommonGuidelines.pdf

Jonassen, D. H. (2009). Reconciling a human cognitive architecture. In S. T. T. M. Duffy (Ed.), *Constructivist theory applied to instruction: Success or failure?* (pp. 13–33). New York, NY: Routledge.

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.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery,

problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.
Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, *48*(3), 169–183.
Klahr, D. (2010). Coming up for air: But is it oxygen or phlogiston? A response to Taber's review of *Constructivist Instruction: Success or Failure? Education Review*, *13*(13). Retrieved from http://www.edrev.info/essays/v13n13.pdf
Ladewski, B. (1994). A middle grade science teacher's emerging understanding of project-based instruction. *The Elementary School Journal*, *94*(5), 499–515.
Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
Lin, H.-T., & Fishman, B. J. (2006). Exploring the relationship between teachers' curriculum enactment experience and their understanding of underlying unit structures. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 432–438). Mahwah, NJ: Lawrence Erlbaum.
Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (2003). *Designing professional development for teachers of science and mathematics* (2nd ed.). Thousand Oaks, CA: Corwin Press.
Lynch, S. (2001). "Science for all" is not equal to "one size fits all": Linguistic and cultural diversity and science education reform. *Journal of Research in Science Teaching*, *38*(5), 622–627.
Lynch, S. J., Pyke, C., & Grafton, B. H. (2012). A retrospective view of a study of middle school science curriculum materials: Implementation, scale-up, and sustainability in a changing policy environment. *Journal of Research in Science Teaching*, *49*(3), 305–332.
MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Lawrence Erlbaum.
MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593–614.
Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*(1), 14–19.
McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*, 233–268.
McNeill, K. L., & Krajcik, J. (2007). Instructional strategies to support students writing scientific explanations. In J. Luft, J. Gess-Newsome, & R. Bell (Eds.), *Science as inquiry in the secondary setting*. Washington, DC: National Science Foundation.
National Research Council. (1999). *Designing mathematics or science curriculum programs: A guide for using mathematics and science education standards*. Washington, DC: National Academy Press.
National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press.
National Research Council. (2001). *Classroom assessment and the national science education standards*. Washington DC: National Academy Press.
National Research Council. (2002). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education*. Washington, DC: National Academy Press.
National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academy Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academy Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.

Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in Earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, *48*(4), 996–1025.

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed teaching observation protocol (RTOP)* (ACEPT Report no. IN-003). Retrieved from https://mathed.asu.edu/instruments/rtop/RTOP_Reference_Manual.pdf

Pituch, K. A., Murphy, D. L., & Tate, R. L. (2010). Three-level models for indirect effects in school- and class-randomized experiments in education. *The Journal of Experimental Education*, *78*(1), 60–95.

Powell, J. C., & Anderson, R. D. (2002). Changing teachers' practice: Curriculum materials and science education reform in the USA. *Studies in Science Education*, *37*, 107–135.

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

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., Martinez, A., Bloom, H., & Hill, C. (2011). Optimal Design Plus Empirical Evidence (Version 3.0) [Computer software]. Available from www.wtgrantfoundation.org

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

Rutherford, F. J. (2000). Coherence in high school science. *Making sense of integrated science: A guide for high schools*. Colorado Springs, CO: BSCS.

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, *50*(5), 1020–1049.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., . . . Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245–253.

Schmidt, W., Houang, R., & Cogan, L. (2002). A coherent curriculum: A case of mathematics. *American Educator*, *26*(2), 10–26.

Schmidt, W. H., McKnight, C., Houang, R., Wang, H.-C., Wiley, D., Cogan, L., & Wolfe, R. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco, CA: Jossey-Bass.

Schmidt, W., McKnight, C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics*. Boston, MA: Kluwer.

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. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education*, *13*(3), 221–245.

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

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*(1), 5–14.

Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, *99*(4), 808–820.

Tobias, S., & Duffy, T. (2009). *Constructivist instruction: Success or failure?* New York, NY: Routledge.

Usiskin, Z. (1985). We need another revolution in secondary school mathematics. In C. R. Hersch (Ed.), *The secondary school mathematics: 1985 yearbook of the National Council of Teachers of Mathematics* (pp. 1–21). Reston, VA: NCTM.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Manuscript received September 11, 2013
Final revision received January 21, 2015
Accepted March 10, 2015