Received: 10 May 2017 | Revised: 24 October 2017 | Accepted: 30 October 2017
DOI: 10.1002/sce.21328

WILEY

RESEARCH ARTICLE

What are critical features of science curriculum materials that impact student and teacher outcomes?

Natalie Pareja Roblin[1,2] | **Christian Schunn**[1] | **Susan McKenney**[3]

[1]Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA

[2]Educational Studies, Ghent University, Ghent, BE

[3]Department of Teacher Professional Development, University of Twente, Enschede, The Netherlands

Correspondence
Susan McKenney, Department of Teacher Professional Development, University of Twente, Enschede, 7500AE, The Netherlands.
Email: susan.mckenney@utwente.nl

Abstract

Large investments are made in curriculum materials with the goal of supporting science education reform. However, relatively little evidence is available about what features of curriculum materials really matter to impact student and teacher learning. To address this need, the current study examined curriculum features associated with student and teacher outcomes. We reviewed sample curriculum materials and documentation reporting on the instructional outcomes of 51 research-based K–12 science curriculum materials. Our findings reveal that teacher supports, rather than student supports, had positive impacts on both student and teacher outcomes. Specifically, positive student outcomes were associated with curriculum materials with a larger scope and with materials that provide teachers with information about students' ideas and recommended instructional strategies. Positive teacher outcomes were associated with the presence of information about targeted standards and recommended instructional strategies. Relatively fewer studies reported on teacher outcomes, and evidence about other dimensions of curriculum materials impact (e.g., spread, sustainability) was difficult to find. Overall, these results reveal the broad importance of embedding teacher supports and ensuring sufficient scope of content across coordinated curriculum units to support the development of conceptual understanding over time. Implications for the design of new curriculum materials and further research are discussed.

KEYWORDS
curriculum materials, science education, student learning, systematic review, teacher learning

1 | INTRODUCTION

Science education has been the focus of recent reform efforts around the world. In the United States, the *Framework for K–12 Science Education* (National Research Council [NRC], 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013) set an ambitious vision for science learning that integrates science and engineering practices, disciplinary core ideas, and crosscutting concepts—thereby moving away from previous approaches focused on learning content and inquiry in isolation (Krajcik, Codere, Dahsah, Bayer, & Mun, 2014). Directly used by both students and teachers, new curriculum materials can provide one crucial form of support for meeting this vision of science learning (Carlson, Davis, & Buxton, 2014).

Curriculum materials have long been put forward as a vehicle of reform since they provide targeted and detailed support for the enactment of specific classroom practices (Brown, 2009; Carlson et al., 2014; Remillard, Harris, & Agodini, 2014). Prior research has revealed many features of curriculum materials that can afford student and teacher learning. However, within the wide range of potentially relevant curriculum features identified in the literature, it remains unclear how specific features compare to each other in terms of impact on instructional outcomes. As a consequence, curriculum developers are left with little guidance as to which features should be prioritized when designing materials that meet the needs of diverse students, teachers, and school contexts. The present study seeks to address this gap in the literature by identifying critical curriculum features empirically associated with student and teacher outcomes across a large sample of research-based K–12 science curriculum materials. In the following sections, we first discuss the *instructional outcomes* that might result from science curriculum materials and then describe various *curriculum features* that have been observed to influence student and teacher outcomes.

2 | THEORETICAL BACKGROUND

2.1 | Outcomes of science curriculum materials

Curriculum materials can be generally defined as resources designed for use by teachers in the classroom to guide their instruction, including textbooks, supplementary units or modules, and instructional media (Remillard et al., 2014). Because of their unique position within the classroom, these materials influence what teachers and students do on a daily basis (Ball & Cohen, 1996; Brown, 2009), and hence can have a significant impact on both student and teacher learning.

Regarding students, many studies show that science curriculum materials can have positive effects on student learning, including gains in students' attitudes and motivation toward science (e.g., Häussler & Hoffmann, 2002; White & Frederiksen, 2000), gains in their understanding of key science concepts (e.g., Harris et al., 2015; Sadler, Romine, Menon, Ferdig, & Annetta, 2015), and gains in their abilities to engage in science practices. The latter include a range of abilities such as the development and use of models (e.g., Schwarz et al., 1986), the planning and carrying out of investigations (e.g., Rivet & Krajcik, 2004), the analysis and interpretation of data (e.g., Marx et al., 2004), the construction of scientific explanations (e.g., McNeill, Lizotte, Krajcik, & Marx, 2006), the engagement in argument from evidence (e.g., Berland & Reiser, 2009), and the evaluation and communication of information (e.g., Cervetti, Barber, Dorph, Pearson, & Goldschmidt, 2012). More recently, there are also some studies that indicate that curriculum materials can significantly enhance students' views about the nature of science (e.g., Akerson, Nargund-Joshi, Weiland, Pongsanon, & Avsar, 2014; Meyer & Crawford, 2015), defined as an understanding of the values and assumptions that are inherent in the development and application of scientific knowledge (Lederman, Antink, & Bartos, 2014). Further, well-designed curriculum materials may also contribute to the development of a broader set of 21st century skills, such as problem solving (e.g., Lester et al., 2014) and students' identities as citizen scientists (e.g., Gaydos & Squire, 2012); however, the impact of science curriculum materials on this broader set of competencies has been rarely investigated.

Beyond the potential effects on student learning, curriculum materials may additionally impact various teacher outcomes. Specifically, research suggests that curriculum materials may play a key role in supporting the development of teachers' science content knowledge (Donna & Hick, 2017; Ellins et al., 2014) and pedagogical content knowledge (PCK; Beyer & Davis, 2012; Marco-Bujosa, McNeill, González-Howard, & Loper, 2017; Schneider & Krajcik, 2002), which is the knowledge that teachers use in transforming subject matter knowledge into forms that are comprehensible for students (Shulman, 2013), such as knowledge of science-specific instructional strategies and of students' thinking about science ideas (Park, Jang, Chen, & Jung, 2011; Schneider & Krajcik, 2002). In addition, curriculum materials may also shape teachers' instructional practices (Beyer & Davis, 2012; Beyer, Delgado, Davis, & Krajcik, 2009; Wyner, 2013), defined as the pedagogical approaches used by teachers to support student learning (Hayes & Trexler, 2016). Finally, curriculum materials may have an influence on teachers' beliefs about science teaching and learning, the nature of science, or about themselves as knowers of science (Dias, Eick, & Brantley-Dias, 2011; Wyner, 2013), as well as on their self-efficacy (Ellins et al., 2014; Sinha et al., 2010). Teacher self-efficacy is defined as teachers' beliefs in their own ability to plan, organize, and carry out activities that are required to attain given educational goals (Skaalvik & Skaalvik, 2010).

It should be noted that the different student and teacher outcomes described above may influence and be influenced by each other. For example, changes in teachers' PCK can potentially impact instructional practices, and this may in turn have a positive influence on student outcomes. Similarly, evidence of improvements in student learning may influence changes in teacher beliefs (cf. Guskey, 2002).

2.2 | Features of science curriculum materials that may impact outcomes

Various features of curriculum materials can play an important role in affording students' and teachers' learning opportunities (Cohen & Ball, 1999; Remillard, 2005). In the present study, we focus on *three sets of curriculum features* that are regularly found across diverse types of curriculum materials (e.g., textbooks, simulation software) and that have been observed to influence instructional outcomes, namely supports for student learning, teacher supports, and scope. Below we provide a description of each set of features and discuss available evidence of their potential impact on student and teacher learning. The specific hypotheses (H1 through H3) relevant to each set of curriculum features are described in the next section (see Figure 1).

Supports for student learning (H1). Curriculum materials often include different types of supports that enable students to accomplish tasks that otherwise might be out of their reach (Reiser, 2004). Some of these supports are gradually withdrawn over time so as to allow students greater responsibility over their own learning (McNeill et al., 2006), while other supports may be more permanent (Pea, 2004). Three different types of supports for student learning are commonly identified in the literature: cognitive, accommodation, and motivational supports. Cognitive supports assist students by highlighting key ideas or relationships among relevant concepts (e.g., outlines, concept maps), by providing additional information for completing a complex task (e.g., worked examples, hints), and/or by giving them opportunities to assess what they know and how their actions change as they learn (e.g., automated feedback, reflection prompts). Accommodation supports provide specific aids to meet the needs of diverse students, such different ability levels (e.g., differentiated tasks), English language support (e.g., access to content in both native and second language, audio support), or special needs (e.g., graphic organizers, adjustable text size). Finally, motivational supports enhance students' interest and engagement in the learning task by, among others, increasing their perceptions of the intrinsic value of the task (e.g., use of driving questions that intrigue students), promoting mastery goals (e.g., encourage students to establish short-term goals in addition to long-term goals), promoting expectancy for success (e.g., use peer modeling to show that the task is neither too difficult nor to easy), and/or promoting student autonomy (e.g., use of noncontrolling language) (Belland, Kim, & Hannafin, 2013).

Various studies have shown positive impacts on student outcomes for cognitive (e.g., Belland, Walker, Olsen, & Leary, 2015; McNeill et al., 2006; White & Frederiksen, 2013) as well as accommodation supports (e.g., Clark, Touchman, Martinez-Garza, Ramirez-Marin, & Drews, 2012; Knight, Spooner, Browder, Smith, & Wood, 2013). However, the motivational dimensions of student supports have received much less attention in the research literature and are rarely included in the design of instructional materials (cf. Belland et al., 2013). Therefore, in the present study we decided to focus only on cognitive and accommodation supports.

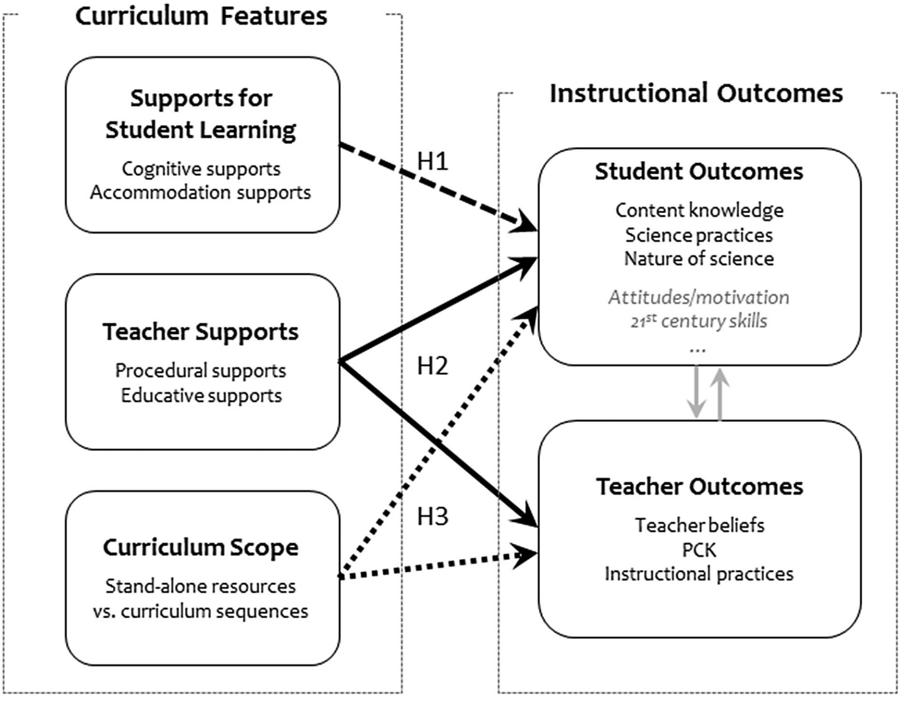

FIGURE 1 Theoretical model

Teacher supports (H2). The effective implementation of innovative curriculum materials can be facilitated by written supports embedded in teacher materials. Here we distinguish between procedural and educative teacher supports. Procedural supports assist teachers in the implementation of the curriculum materials by providing them with organizational information (e.g., time and resources required, advanced preparation), with recommended instructional strategies to productively support student learning (e.g., tips for leading whole-class discussions), and/or with examples of how others have adapted the curriculum materials to their specific classroom situations (e.g., lesson plans and/or worksheets created by other teachers). Research suggests that such procedural supports are necessary to render the materials practical for everyday use (Janssen, Westbroek, Doyle, & Van Driel, 2013).

While some procedural supports may indirectly contribute to teacher learning, educative supports are expressly designed to provide direct opportunities for teacher learning (Davis & Krajcik, 2005; Grossman & Thompson, 2008). Specifically, educative supports aim to (1) help teachers anticipate student thinking and misconceptions, (2) support teachers' learning of the subject matter, (3) help teachers consider ways to relate units during the year, (4) make visible the rationale behind particular design decisions, and (5) promote teachers' capacity to implement and adapt the curriculum materials (Davis & Krajcik, 2005). There is growing evidence about the positive impact of educative supports on the development of teachers' PCK (e.g., Beyer & Davis, 2012; Marco-Bujosa et al., 2017; Schneider & Krajcik, 2002) and on their instructional practices (e.g., Arias, Davis, Marino, Kademian, & Palincsar, 2016; Cervetti, Kulikowich, & Bravo, 2015), as well as on how these may, in turn, influence student learning (Arias, Smith, Davis, Marino, & Palincsar, 2017; Bismack, Arias, Davis, & Palincsar, 2015).

Scope (H3). Scope refers to the breadth of content covered by the curriculum materials in a given subject area (Grossman & Thompson, 2008; NRC, 2012), which may range from stand-alone curriculum units or tools targeted toward specific content that teachers can flexibly integrate into their instruction, to comprehensive curriculum materials covering one or multiple years of instruction (Carlson & Anderson, 2002). Research suggests that curriculum materials with a larger scope (e.g., a 3-year curriculum covering all key contents for middle school science or a sequence of three coordinated curriculum units about energy) may help students build their understanding over time and develop increasingly sophisticated ideas (Fortus, Sutherland Adams, Krajcik, & Reiser, 2015; Stevens, Delgado, & Krajcik, 2015), thereby generating a positive impact on student learning. Moreover, curriculum materials with a larger scope may provide explicit pedagogical guidance and diverse supports for teachers, and hence increase opportunities for teacher learning (Grossman & Thompson, 2008).

3 | STUDY PURPOSE AND HYPOTHESES

As described above, the specific features of curriculum materials can have an important impact on student and teacher outcomes. A number of studies and reviews have enumerated and argued for the importance of many curriculum features that could each contribute to student and teacher learning (cf., Kesidou & Roseman, 2002). While some core features have received clear empirical support (Beyer et al., 2009; Fortus et al., 2015; McNeill et al., 2006; Schneider & Krajcik, 2002), their impacts are potentially limited to particular contents, contexts, or types of curriculum materials. Further, each of the features is potentially expensive to integrate in robust and high-quality ways, and too many features might prove overwhelming. Little is known about how different curriculum features compare to each other in terms of impact on instructional outcomes. As a consequence, curriculum developers are left with little guidance concerning which features should be prioritized when designing materials that meet the needs of diverse students, teachers, and school contexts. The present study seeks to address this gap in the literature by identifying key curriculum features empirically associated with positive student and teacher outcomes across a large and diverse sample of research-based K–12 science curriculum materials. Based on the research evidence discussed in the preceding section, the following three hypotheses were formulated:

1. (H1) Cognitive and accommodation supports for students embedded in curriculum materials are associated with positive student outcomes.
2. (H2) Procedural and educative teacher supports embedded in curriculum materials are associated with positive teacher and student outcomes.
3. (H3) Curriculum materials with a larger scope are more likely to yield positive student and teacher outcomes.

4 | METHOD

4.1 | Sample

To determine the features that influence the impact of science curriculum materials on student and teacher outcomes, a broad sample of research-based curriculum materials was obtained, focusing on those funded by the National Science Foundation (NSF) and the Institute of Education Sciences (IES). The sample was limited to projects awarded between 2001 and 2010 to ensure access to project documentation. For projects awarded before 2001, it could be difficult to obtain information about the materials' features, while for projects awarded after 2010 dissemination of results may have not occurred by the time of document collection and analysis (2015–2016). Although the curriculum materials developed by these projects preceded the release of NGSS, most of them integrated science content and practices since this was typically asked for in the requests for proposals of the funding agencies.

We obtained and analyzed documentation and sample curriculum materials from 51 projects concerned with the design of research-based K–12 science curriculum materials for classroom use. The identification and selection of these projects followed a six-step procedure (see Table 1 for an overview). NSF and IES grant awards concerned with curriculum design were sought in the official databases of each funding agency (step 1). To be deemed relevant, awards had to meet the following inclusion criteria: target mainstream K–12 science education, have curriculum design as an important goal, and focus on curriculum materials for classroom use (step 2). Different awards linked to the same project were then merged (step 3). A pilot phase of debugging the coding process with 13 randomly selected projects

TABLE 1 Identification of relevant projects

Step	Description	Total
1	Search hits for NSF and IES *awards*	1,301
2	Identification of *awards* concerned with curriculum materials for K–12 science classroom use	162
3	Identification of *projects* concerned with the design of K–12 science curriculum materials for classroom use (after merging related awards)	146
4	Random sample of projects	83
5	Description and verification of key features of the resulting curriculum materials	69
6	Identification of projects with available documentation about learner and teacher outcomes	51

established that the coding was very labor intensive. Therefore, a subset of projects was randomly selected to make the process of in-depth analysis of curriculum features and project outcomes manageable (step 4). The selection process was implemented at the level of principal investigator rather than at level of project to ensure efficient use of principal investigator time for member-check interviews. When a principal investigator received multiple awards within our 10-year time window, all the awards were included. This resulted in 57% of the total sample.

Only projects for which member-check interviews were completed to verify the features of the curriculum materials were kept for further analyses (step 5), so as to ensure both accuracy of the coding and access to project documentation provided by the principal investigator. Finally, projects with no available documentation reporting on student or teacher outcomes were excluded from further analyses (step 6), resulting in a final sample of 51 projects. Chi-square tests revealed no statistical significant differences (using an alpha level of $p < .05$) between the projects included in the sample ($n = 51$) and those excluded ($n = 95$) with regard to metadata collected on all projects (i.e., award start year period, target grade level, and science discipline).

Most of the sampled projects targeted middle (33%) and high (29%) school, while 16% targeted elementary school. A number of projects (22%) targeted multiple grade bands. Regarding main science disciplines, over half of the projects developed curriculum materials targeting a single science discipline (life sciences: 18%; physical sciences: 16%; environmental sciences: 12%; earth and space sciences: 8%). The remaining projects developed curriculum materials either for more than one science discipline (25%), or for the explicit integration of two or more science disciplines (e.g., a unit in biochemistry; 22%). The format of the curriculum materials also varied between projects. Fifty-three percent of the projects developed mostly printed materials (e.g., books, worksheets), whereas 47% developed primarily digital materials (e.g., simulations, websites, educational games).

4.2 | Measures

Predictor variables. A detailed coding scheme was designed to characterize the curriculum materials produced by the projects in the sample. The instrument development was informed by existing literature about key features of science curriculum materials associated with positive student and teacher outcomes (see the section Theoretical background). To ensure that the coding scheme could be applied to curriculum materials with diverse formats (e.g., textbooks, simulation software) and target audiences (e.g., elementary, secondary school), the selected features had to be (1) broadly representative across diverse types of curriculum materials and (2) associated with the types of instructional outcomes typically addressed in evaluation studies. For example, although most curriculum materials are likely to include features aimed at triggering students' motivation and interest in science, the availability of such features would be difficult to code as these are deeply embedded in the materials (e.g., use of noncontrolling language). Further, a preliminary review of project documentation revealed that motivational outcomes were not measured on a regular basis, whereas cognitive or accommodation support outcomes were.

TABLE 2 Supports for student learning embedded in curriculum materials

Code	Description	Examples
Cognitive supports (*no* = 0; *yes* = 1)	Supports intended to assist learners by highlighting key ideas or relationships among concepts, providing additional information for completing a task, and/or giving opportunities to assess what they know	- Concept maps/outlines - Worked examples - Hints - Reflection prompts - Automated feedback
Accommodation supports (*no* = 0; *yes* = 1)	Specific supports for English language learners, learners with different ability levels and/or learners with special needs	- Translations to other languages - Text to speech - Graphic organizers - Adjustable text size - Differentiated tasks for students with different ability levels

To assess content validity, the coding scheme was extensively discussed with two curriculum developers with more than 10 years of experience in curriculum design, and who had produced highly effective curriculum materials (as demonstrated by the outcomes of rigorous evaluation studies and a prize for excellence in educational design). The curriculum experts were requested to comment on the extent to which the codes were both theoretically relevant and sufficiently well represented across diverse types of curriculum materials. Refinements were made based on their input. The final version of the instrument consisted of three sets of curriculum features, these related to supports for student learning (see Table 2), teacher supports (see Table 3), and scope (see Table 4).

Outcome variables. An instrument was also developed to describe projects' instructional outcomes. First, student and teacher outcomes were initially operationalized based on the research literature about the impact of science curriculum materials presented in the section Theoretical background above. Then, these outcomes were refined through inductive analyses of a small sample of project documents to determine what outcome information (1) could be distinguished across heterogeneous study designs and reporting methods and (2) was regularly available. For example, although the literature frequently distinguishes PCK from pedagogical knowledge, studies in this data set often did not report measures of pedagogical knowledge alone. Further, while the science education literature often mentions the importance of student motivation, the coded studies examined motivation relatively rarely.

To assess the content validity of the instrument, feedback was sought from (science) education researchers and curriculum designers. Initially, individual telephone interviews with six (science) education researchers were conducted to gather their feedback on the extent to which the outcome variables identified are broadly representative of student and teacher outcomes in science education. Then, a revised version of the instrument was discussed with a group of curriculum designers to assess its overall clarity and relevance across the projects investigated. Tables 5 and 6 contain an overview of the different outcomes studied, the ways they were operationalized, and examples of outcome evidence from project documentation for each variable.

4.3 | Procedures

Description of key features of the curriculum materials. Publicly available samples of the curriculum materials as well as project publications describing them (e.g., evaluation reports, journal articles, conference papers) were gathered. Information from all available sources was equally weighed and triangulated to ensure coding accuracy. In the case of inconsistencies across sources, principal investigators were probed during member-check interviews.

TABLE 3 Teacher supports embedded in curriculum materials

Code	Description	Examples
Procedural supports (*no* = 0; *yes* = 1)		
Organizational information	Information about time and resources required to effectively implement the curriculum materials	- Number of class periods required - List of materials and equipment - Safety guidelines - Advanced preparation
Instructional strategies	Specification of recommended pedagogical approaches to support student learning	- Pedagogical suggestions to productively support student learning - Tips for leading whole-class discussions - Recommendations for how to integrate the activities or tools in the curriculum
Access to variations created by others	Examples of how other teachers have adapted the curriculum materials to their specific classroom situations (beyond exchanges during professional development workshops) and/or access to a platform where teachers can share their adapted lesson/activity plans	- Lesson plans, worksheets and/or activities created by teachers available for download from project website - Platform where teachers can share their modified lesson plans, worksheets and/or experiences with others
Educative supports (*no* = 0; *yes* = 1)		
Unit goals	Brief statements of what students will be able to learn, which could help teachers to link different units	- Overview of learning goals for each curricular unit, lesson or activity
Alignment to standards	Explicit references to the standards addressed in the curriculum materials	- List of national and/or State content standards addressed in the curriculum materials
Information about students' ideas	Information that could help teachers to anticipate student thinking and (alternative) conceptions	- Overview of prerequisite knowledge (i.e., what students should already know and/or be able to do). - Overview of common misconceptions (i.e., misunderstandings and/or or difficulties students have shown to have with a particular content/activity in the materials)
Subject matter support	Information that could assist teachers' learning of the science content addressed in the materials beyond the level of understanding required by students	- Description of key science concepts addressed in the curriculum materials - Links to websites and/or resources with additional information about the key science ideas addressed in the curriculum materials

Coding of the key features of the curriculum materials produced by each project resulted in individual project profiles which were verified with principal investigators in a telephone interview to check the reliability of the coding and collect information for missing fields. Member-check interviews were completed for 69 (83%) of the projects selected at step 4 of the sampling procedure (see Table 1); a few principal investigators had retired or did not return e-mails. During the interviews, the coding of each curriculum feature was discussed until principal investigators were satisfied with its characterization. Disagreements between coders and principal investigators were very rare, and conversations focused mostly on adding information for missing fields. In the few cases that coding did not accurately reflect a

TABLE 4 Scope of the curriculum materials

Code	Description	Examples
Stand-alone curriculum resources (= 0)	Stand-alone curriculum units, activities, or tools structured around a specific topic, and which may be used to supplement or replace parts of existing curricula. Even though multiple curriculum units or activities may be developed by a project, these do not need to follow a particular sequence or build on each other	- Interactive models of atomic structure - A 6-week curriculum unit about photosynthesis
Curriculum sequences (= 1)	Series of two or more consecutive curriculum units that build on each other, covering a larger range of content/key understandings for a specific grade level(s) and science discipline(s)	- A sequence of three consecutive curriculum units on plate tectonics designed to scaffold students' technical skills, analytical abilities, and scientific content knowledge, as well as increase students' understandings about the nature of science - Three-year curriculum that introduces students to all of the core concepts in inquiry, physical sciences, life sciences, and earth–space sciences found in the national standards for Grades 9–12.

feature of the curriculum materials, it was because of revisions made during the development process (e.g., a project initially targeting both middle and high school that eventually ended up developing materials for middle school only). Here, principal investigators provided recent documentation about the latest version of the materials and explained the rationale for the changes made. As noted in the sampling procedure, projects for which member-check interviews could not be completed were excluded from further analyses.

Identification of relevant documentation reporting on projects' instructional outcomes. Studies and reports produced by the 51 sampled projects were sought on project websites, principal investigator's personal websites, and scientific databases (Google Scholar and Web of Science). Keywords used to search for project documentation in scientific databases included award number, project title, and/or (co-) principal investigator's name. In a member-check interview, principal investigators were asked to go over the search results and identify additional publications that could be relevant to our analyses. This resulted in the identification of over 500 references to publications and project documents. We were unable to obtain the full text for 43 of these references, the majority of which (53%) were conference presentations.

Available publications and documents were then screened to determine relevance to our study. To be deemed relevant, documents had to report on the effects of the curriculum materials on at least one of the outcome variables described in the previous section (see Tables 5 and 6). When multiple documents reported results from the same study, peer-reviewed and/or most recently published sources were prioritized, and duplicates were removed. To avoid redundancy, doctoral dissertations were not included unless they were the only source of evidence available for a project. Finally, only documents concerned with the outcomes of the curriculum materials developed by the specific award(s) in the portfolio review were deemed relevant; documents reporting on the outcomes of subsequent development work (e.g., addition of new units or revisions to the curriculum materials) were not eligible since the key features of the curriculum materials could have changed. However, documentation from subsequent awards specifically concerned with the investigation of the effects of the previously developed curriculum materials was deemed relevant, as long as no changes were made to the curriculum materials.

Coding of projects' instructional outcomes. Available project documentation was reviewed to describe student and teacher outcomes for each project. When a document reported multiple studies, results for each study were coded separately. Moreover, when multiple versions of the curriculum materials were evaluated (e.g., a pilot study followed by a field trial of a revised curriculum unit), only the effects of the latest version were used in the aggregation of project scores (see also description of aggregative scores below).

TABLE 5 Operationalization of coded student outcomes along with examples of evidence on each

Outcome	Definition	Examples of outcomes evidence
Students' understanding of *science content*	Gains in students' understanding of specific science content (e.g., force, motion, chemical reactions)	"The paired-sample test for science content understanding for all students indicated statistically significant gains with a large effect size. Specifically, the average test score increased from 65% on the pretest to 94% on the posttest. The independent samples Mann–Whitney test showed no difference in the distribution of scores for science content understanding across gender. As well, no differences were noted in a fall versus a spring curriculum implementation or across schools."
Students' ability to engage in the *science practices*	Gains in students' ability to engage in the process of doing science, including their ability to (1) ask questions and define problems, (2) develop and use models, (3) plan and carry out investigations, (4) analyze and interpret data, (5) construct explanations and design solutions, (6) engage in argument from evidence, and/or (7) obtain, evaluate, and communicate information	"Scientific reasoning scores for all students indicated statistically significant gains with large effect sizes. The average scores for all students increased from 20 % on the pre-test to 47% on the post-test—a 135% change. On the pretest, the majority of student responses included only a claim and one piece of evidence with no reasoning provided. On the posttest, the majority of students' science explanations included a claim, evidence, and reasoning."
Students' understanding of the *nature of science* (*NOS*)	Gains in students' understanding of the values and assumptions that are inherent in the development and application of scientific knowledge, including understanding that (1) scientific investigations use a variety of methods, (2) scientific knowledge is based on empirical evidence, (3) scientific knowledge is open to revision in light of new evidence, (4) scientific models, laws, mechanisms, and theories explain natural phenomena, (5) science is a way of knowing, (6) scientific knowledge assumes an order and consistency in natural systems, (7) science is a human endeavor, and (8) science addresses questions about the natural and material world.	"Our data show that students also demonstrated significant learning gains in understandings about NOS. For the VNOS test questions, students scored an average of 1.39 points on the pretest (SD = 0.38; SEM = 0.08) and 2.00 points on the posttest (SD = 0.24; SEM = 0.06). These scores were then analyzed using a paired-sample t-test and results showed a significant difference in scores across the pre- and post-measures ($\alpha < .01$), indicating student learning."

Given that it was often impossible to assess the impact of curriculum materials based only on the raw data presented in appendices or in evaluation reports, no secondary analyses of such raw data were conducted. Similarly, secondary sources reporting indirectly on the effects of the curriculum materials (e.g., an annual report summarizing published studies) were not used as evidence; when a document cited possibly relevant references, the original source was always sought.

Document screening resulted in the identification of 170 documents reporting on student or teacher outcomes. Over half of these documents (64%) were peer-reviewed (e.g., journal articles, edited book chapters, conference proceedings or conference papers). Some documents reported multiple studies, resulting in a total of 227 studies, the majority of which used quantitative (59%) or mixed (34%) methods. To capture sufficient evaluation outcomes (i.e., only 35% of the documents were journal articles) and to avoid possible publication bias (i.e., only making public positive outcomes, or only reporting the significant effects in journals), all sources were used in the analyses regardless of study method or publication type. Follow-up chi-square tests revealed no statistical significant differences (*p*s > .6) between the types of instructional outcomes (i.e., No Gains or Mixed Outcomes; Gains) reported in peer-reviewed vs. non-peer-reviewed sources.=

TABLE 6 Operationalization of coded teacher outcomes along with examples of evidence on each

Outcome	Definition	Examples of outcomes evidence
Teachers' *beliefs about science, science education* and/or *NOS*	Changes in teachers' beliefs about the nature of science, science teaching and learning, and/or about themselves as knowers of science, including espoused beliefs (i.e., self-reported claims about the ways things are or should be) and/or beliefs-in-action (i.e., beliefs that implicitly guide and are inferred from teachers' actions)	"A Wilcoxon match-pairs signed ranks test indicates that post implementation teachers placed a significantly ($p = 0.05$) higher value on integrating human impact into the ecology section of the curriculum than they did prior to implementation."
Teachers' *self-efficacy beliefs*	Gains in teachers' beliefs in their own ability to plan, organize, and carry out activities that are required to attain given educational goals	"Reported confidence in teaching the three main topics was good for nearly all respondents after the workshop […] Both Texas and Mississippi teachers reported increased climate knowledge and confidence due to participation in the workshops and expressed the belief that the workshop components would be equally useful to them in the classroom."
Teachers' science *content knowledge* (CK) and/or *pedagogical content knowledge* (PCK)	CK: Gains in teachers' understanding of the science content addressed in the curriculum materials	"Based on the teachers' pre/post/post-post assessment data, the teachers' science content knowledge related to ecosystems was not enhanced through their participation in the professional development or through their experiences teaching science using the materials. Those data indicate that teachers came into the professional development with very high levels of understanding about ecosystems".
	PCK: Gains in teachers' understanding of how to transform subject matter knowledge into forms that are comprehensible for students	"Overall, the survey responses indicate that all of the teachers increased their understanding of the tools and content specific to the [XX] materials, as well as inquiry approaches for teaching science. Most, but not all, of the teachers also increased their understanding of how students learn science, what scientists do, and how to use formative assessment."
Teachers' *instructional practices*	Changes in teachers' instructional practices as a possible result of the curriculum materials, including instructional strategies, classroom norms, teacher and/or student roles and teacher–student interactions	"Analyses indicated that all teachers showed positive movement toward more reformed-based ways of teaching, including using at least some features of inquiry-based instruction. Several teachers made explicit reference to NOS, at least to some degree. This was an improvement when compared to the Pre-work session videos."

Given the heterogeneous types of studies (e.g., quantitative, mixed, and qualitative), and study designs (e.g., within-group gain data, between-group contrast data) used in evaluations of curriculum materials, we opted to not use effect sizes as the basis for aggregation. Effect sizes are not available in qualitative studies and not comparable across different study designs, and therefore could lead to misinterpretations of the magnitude of the effects observed (cf. Olejnik & Algina, 2000). Consequently, evidence of student and teacher outcomes on a particular outcome variable for each individual study was coded on a three-point scale (1 = No Gains; 2 = Mixed Outcomes; and 3 = Gains). For quantitative or mixed-methods studies, results were coded as No Gains when differences between pre- and posttests were not statistically significant ($p > .05$). Results were coded as Mixed when overall pre- to posttest gains were not statistically significant, but (1) results show evidence that the curriculum materials work for some subgroups (e.g., English

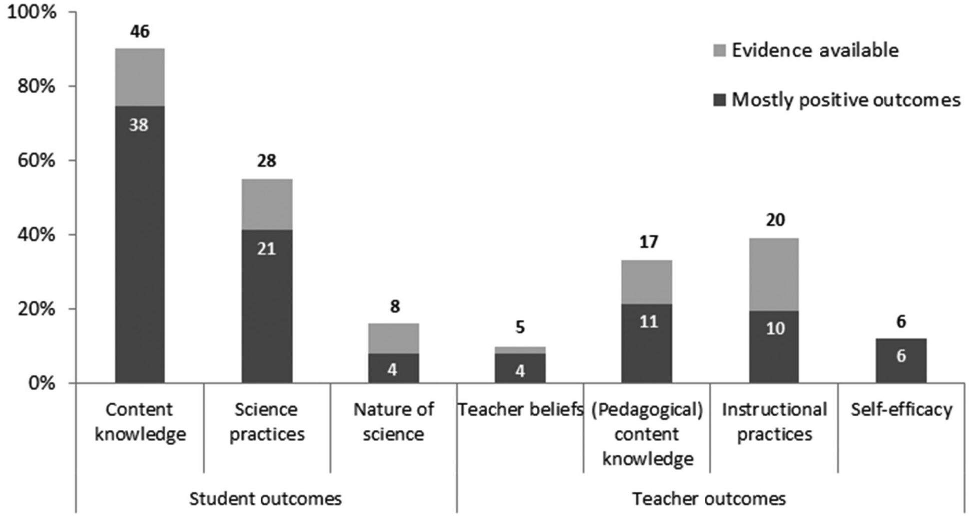

FIGURE 2 Percentage (and number) of projects with available evidence of student and teacher outcomes ($N = 51$). The dark bar presents the number of projects reporting mostly positive outcomes

language learners, high achievers); or (2) results for some subscales were statistically significant (e.g., significant gains on multiple-choice items, but overall no significant differences). Finally, results were coded as Gains when the study showed statistically significant gains (using the alpha level of $p \leq .05$) from pre- to posttest. Quantitative evidence was always coded based on total or composite scores.

For studies reporting qualitative evidence only, results were coded as No Gains when the study claims no changes associated to the use of the curriculum materials and provides quotes to support that claim. Results were coded as Mixed if the study provides evidence of gains for some participants (≤ 50%), but not for others. Finally, results were coded as Gains if the study claims gains for more than half of the participants, provides quotes to support that claim, and no counterevidence.

Reliability. To establish reliability in the coding of instructional outcomes, 26% of the studies reporting on student or teacher outcomes ($N = 60$) were reviewed by two independent researchers, and differences in judgments discussed until acceptable levels of reliability were obtained (student outcomes, $\kappa = 0.76$; teacher outcomes, $\kappa = 0.70$).

A second round of member-check interviews was conducted, this time to verify the outcomes coding. Member-checks were completed for 45 projects in the sample (88%); a few principal investigators were not available or did not return e-mails. During the member-check interviews, which lasted between 60 and 90 minutes, principal investigators were asked to verify if the description of their project's outcomes was accurate and to provide access to additional documentation that could be relevant for our analyses. To facilitate this process, a few days prior to the interview principal investigators were provided with an excel file containing (1) a list of all the documentation found for the project, and of those documents included in the analyses; (2) an overview of the project's student and teacher outcomes; and (3) an overview of the evidence used to code the project's instructional outcomes. New documents provided by the principal investigators were first screened for relevance and then coded using the same procedure described above.

Aggregative scores. Once all individual study codes were verified, two types of aggregate scores were created. First, individual study scores were aggregated at project level for each of the student and teacher outcome variables separately (0 = No Gains or Mixed Outcomes; 1 = 75% or more studies show Gains). These aggregate scores allowed us to address issues of distributions and data availability of individual outcome variables. The 75% threshold was used because there was a positive skew to the reported outcomes. As illustrated in Figure 2, the amount of evidence available about student and teacher outcomes varied largely, with some variables (e.g., students' understanding of science content, students' ability to engage in the science practices) being more commonly investigated than others (e.g., students' understanding of the nature of science, teacher beliefs, teachers' self-efficacy). Moreover, when evidence was available, it was often evidence of positive or mixed outcomes, with relatively fewer studies reporting no significant outcomes.

Second, individual study scores reporting on the effects of the latest version of the curriculum materials for all student outcome variables were aggregated into an overall project score (0 = No Gains or Mixed Outcomes; 1 = 75% or more studies show Gains). The same was done for teacher outcome scores. This produced aggregate scores that allowed us to examine what curriculum features are critical to impact each of these two key instructional outcomes. We also examined curriculum features associated with student outcomes on content knowledge alone, since this was the variable with most evidence available. However, our main analyses focused on aggregated student outcomes because these better represent the integration of the three dimensions of student learning typically targeted by research-based science curriculum materials.

4.4 | Data analysis

Descriptive statistics were used to characterize the curriculum materials developed by the projects included in the sample and to describe patterns in the available evidence of student and teacher outcomes. To have enough power to support statistical analyses, all grade bands were aggregated together. Chi-square tests revealed no significant associations between target-grade band (i.e., elementary school vs. secondary school curriculum) and student ($p = .93$, two-tailed) or teacher ($p = .95$, two-tailed) outcomes. Associations between curriculum features and project outcomes were then assessed using one-tailed chi-square tests. A one-tailed test was chosen because all the hypotheses are directed, predicting positive outcomes when the focal curriculum features are available in the materials.

To ascertain the magnitude of the effect, phi was computed. Phi is a measure of the degree of association between two binary variables. Its value ranges from −1.0 (strong negative association) to +1.0 (strong positive association), with values above ±0.3 generally considered as medium effect and above ±0.5 as strong effect (Cohen, 1988). Finally, follow-up logistic regression analyses were conducted to uncover which among the correlated curriculum features had a significant relationship with student and/or teacher outcomes, when controlling for multiple features at once (i.e., to address possible confounds).

4.5 | Limitations

Several limitations of the study bear mention. First, the differential availability of outcomes evidence may be correlated with key curriculum features, thereby introducing bias. To examine this possibility, we conducted a correlational analysis between key curriculum features and availability of student and teacher outcomes evidence. Results revealed that availability of student outcomes evidence was not significantly associated with any of the studied curriculum features. For teacher outcomes, although two features (cognitive supports for students, $phi = .–.38$; $p < .01$; and examples of variations created by other teachers, $phi = .39$; $p < .01$) were significantly associated with availability of outcomes evidence, none of these turned out to be significantly associated with positive teacher outcomes.

Second, our review included studies with different methodologies (e.g., quantitative, mixed, qualitative) and study designs (e.g., single-group pre–posttest, control-group pre–posttest), thereby introducing noise into the analysis process as is commonly the case in systematic review studies. A related limitation pertains to the large diversity of assessment instruments used by the various studies to measure student and teacher outcomes (e.g., assessments directly aligned with the curriculum materials, standardized tests). Even though the instruments targeted the same constructs (e.g., students' ability to construct explanations, teachers' PCK), they are not necessarily equivalent. Moreover, given that the curriculum materials included in the review were developed prior to the release of NGSS, most assessment instruments focused primarily on science content with relatively fewer assessments specifically targeting students' ability to engage in science practices or their understanding of the nature of science. Assessments of teacher outcomes such as changes in PCK or in teacher beliefs were also scant.

Third, our predictor variables involved the presence of curriculum features, and necessarily aggregated variations of each feature as a holistic group (e.g., student supports) to have enough instances per group to support statistical analyses. New empirical studies would be needed to unpack differences in subtypes of curriculum features (e.g., different forms of cognitive supports for students). A related issue involves the quality (rather than just presence) of curriculum features. Given that our sample included curriculum materials with diverse formats (e.g., textbooks, simulations), length (e.g., few class periods, several instructional years), and target audiences (e.g., elementary school, secondary school), our analyses had to necessarily remain at a general level. Having a unique quality measure to assess the scope and depth of each curriculum feature across such large diversity of project types would be very difficult. Further, there would be too few instances of each variation to identify meaningful patterns given the likely confounds in the diverse pool.

Finally, the effects of curriculum materials could also be influenced by other factors such as the availability and quality of professional development opportunities (Garet, Porter, Desimone, Birman, & Yoon, 2001), teachers' prior experience with the curriculum materials (Schneider, 2009), the diverse ways teachers interact with and adapt the curriculum materials (Fogleman, McNeill, & Krajcik, 2011; Remillard, 2005; Schneider, Krajcik, & Blumenfeld, 2005), and the available support structures at the school and district level (Carlson et al., 2014; Harris et al., 2015). Further research is required to understand interactions between these factors and the effects of curriculum material features on student and teacher outcomes.

5 | RESULTS

5.1 | Descriptive results of curriculum materials' features

Table 7 provides an overview of the proportion of projects that include each of the studied curriculum features, as well as bivariate correlations to examine the presence of highly co-occurring features (i.e., possible confounds). As illustrated in the table, most of the correlations are weak to medium, thereby allowing for investigation of independent effects. For variables correlated with each other and also associated with positive instructional outcomes, multiple regressions were conducted.

TABLE 7 Number of projects (*n*), proportion of projects with the curriculum feature (*M*), and bivariate correlations (phi)

			Correlations								
Curriculum features	***n***	***M***	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
(1) Scope	51	0.31		–							
Supports for student learning	50										
(2) Cognitive supports		0.76	.18		–						
(3) Accommodation supports		0.46	.10	**.35***		–					
Procedural teacher supports	51										
(4) Organizational information		0.88	.25	.22	.09	–					
(5) Instructional strategies		0.82	.09	-.02	-.09	**.47****	–				
(6) Variations created by others		0.29	.03	**-.35****	-.08	-.03	.07	–			
Educative teacher supports											
(7) Alignment to standards	51	0.78	.15	-.07	.10	**.40****	.26	.02	–		
(8) Unit(s) goals	51	0.94	.17	.06	.06	**.68****	**.54****	.16	**.27***	–	
(9) Students' ideas	50	0.48	**.37****	.27	.06	**.35****	.24	.07	.03	.24	–
(10) Subject matter support	50	0.76	**.29***	.01	-.15	**.51****	**.35****	.06	**.38****	**.45****	**.35****

$^{**}p<.01$, two-tailed; $^{*}p<.05$, two-tailed, with statistically significant correlations in bold.

TABLE 8 Phi values for the association of curriculum features with positive student and teacher outcomes

Predictor variables	Student outcomes ($n = 50$)	Teacher outcomes ($n = 29$)
Supports for student learning (*no* = 0; *yes* = 1)		
Cognitive supports	−.04	−.08
Accommodation supports	−.09	.03
Teacher supports (*no* = 0; *yes* = 1)		
Procedural supports		
Organizational information	.16	.05
Instructional strategies	**.26***	**.40****
Variations created by others	.05	.19
Educative supports		
Alignment to standards	.07	**.34***
Unit(s) goals	.20	.22
Information about students' ideas	**.39****	−.25
Subject matter support	**.24***	.09
Scope (*stand-alone resources* = 0; *curriculum sequences* = 1)	**.36****	−.02

**p<.01, two-tailed; *p<.05, two-tailed, with statistically significant correlations in bold.

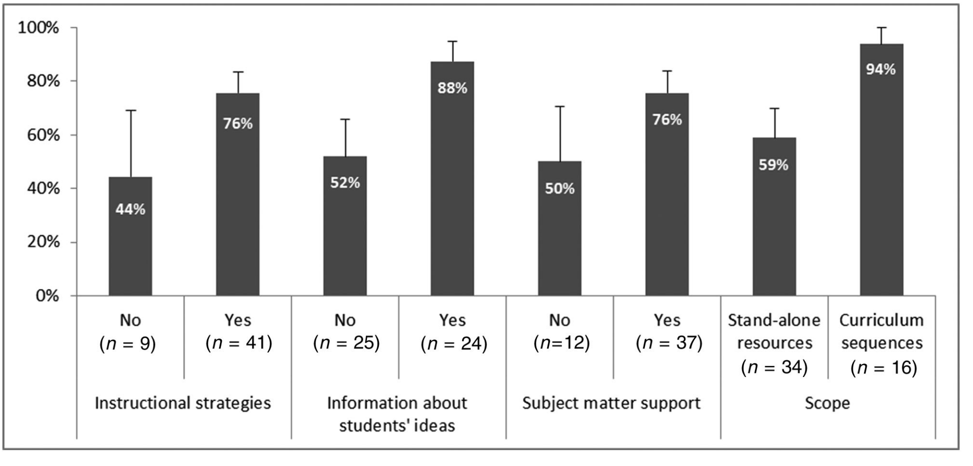

FIGURE 3 Proportion of projects with mostly positive student outcomes (with standard error bars) for projects with and without key curriculum features (with *n*s for each)

5.2 | Curriculum features associated with positive outcomes

In this section, we examine curriculum features associated with positive student and/or teacher outcomes. Note that student and teacher outcomes were only weakly correlated at the project level (*phi* = −.07, p = .72), such that separate analyses at each level are meaningful. Table 8 contains the results of one-tailed chi-square tests, and Figures 3 and 4 provide an overview of those curriculum features that have a significant effect on student and teacher outcomes, respectively. A representative example of each critical curriculum feature drawn from materials with positive outcomes is presented in Table 9.

As the instructional strategies example in Table 9 notes, the materials sometimes contained multiple strategies. This was true in most of the projects with positive outcomes, so this example should not be taken as the definitive recommended form. Other example strategies from projects with positive outcomes include providing teachers

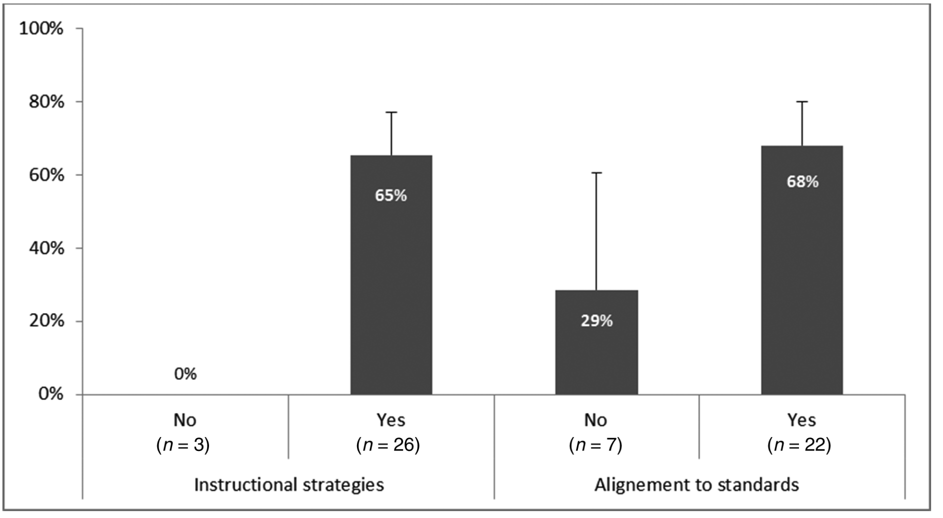

FIGURE 4 Proportion of projects with mostly positive teacher outcomes (with standard error bars) for projects with and without key curriculum features (with *n*s for each)

information on when to withhold answers from students to better support exploration (Kowalski, Van Scotter, Stuhlsatz, & Taylor, 2010), scripts for facilitating whole-class discussions (Lachapelle et al., 2011), and linguistic scaffolding strategies (Meyer & Crawford, 2011). The most detailed descriptions of the instructional strategies are typically found in the teacher guides rather than in journal articles or evaluation reports, which tend to emphasize the outcomes, the design of the student materials, or the professional development provided to the teachers. the included in their projects.

Student outcomes. Four key features of curriculum materials had a weak to medium association with mostly positive student outcomes (see Figure 3). Specifically, curriculum materials that included a description of instructional strategies that teachers can use to facilitate student learning were more likely to yield positive student outcomes. The availability of information about students' ideas (e.g., prerequisite knowledge or alternative conceptions) and of subject matter support was also significantly associated with student outcomes. Further, results revealed a significant association between student outcomes and the scope of the curriculum materials. Materials with a larger scope, such as comprehensive curricula or curriculum sequences, were more likely to yield positive student outcomes.

No significant associations were found between student outcomes on the one hand, and the presence of cognitive or accommodation supports for students on the other hand. Results using only outcomes for content knowledge revealed a very similar pattern, although association with unit goals became stronger ($phi = .34$; $p \leq .01$) and association with subject matter support was weaker ($phi = .11$; $p > .05$).

Teacher outcomes. Two curriculum features had a medium association with mostly positive teacher outcomes: instructional strategies and alignment to standards (see Figure 4). Curriculum materials that provided teachers with a description of instructional strategies were more likely to result in positive teacher outcomes, whereas no positive outcomes were observed for curriculum materials that did not specify instructional strategies. Similarly, curriculum materials that provided teachers with information about the national and/or state standards addressed in the curriculum materials were more likely to result in positive teacher outcomes than were curriculum materials not including such information.

In contrast to our hypothesis, no statistically significant associations were found between teacher outcomes and the scope of the curriculum materials. Teacher outcomes were also not significantly associated with other forms of educative teacher supports, such as the presence of unit goals, information about students' ideas, or subject matter support.

TABLE 9 One representative example of each critical curriculum feature drawn from materials with positive outcomes

Critical Feature	Example from Curriculum Materials
Scope	The *Investigating and Questioning our World through Science and Technology* (IQWST) is a comprehensive 3-year middle school science curriculum that targets core ideas in physics, chemistry, life science, and earth science and explicitly designed to address curriculum coherence. The units are designed to build on one another and extend science ideas and practices, thereby helping students build clear connections across ideas and units. Each lesson begins and ends with an explicit link to the previous and following lesson. Source: Roseman, Fortus, Krajcik, and Reiser (2015).
Alignment to standards	*EarthLabs* is a collection of eight online modules that provides rigorous and engaging earth and environmental science labs. Teachers may use the entire collection as presented or choose specific modules to include in the courses that they teach. The *EarthLabs* Climate modules are aligned with the Climate Literacy Essential Principles. The modules also address the Texas Essential Knowledge and Skills (TEKS) for Science and the 2010 Mississippi Science Framework. These alignments improve educators' ability to understand and use the Climate Literacy Principles, as well as to teach their state standards effectively. Source: Ellins et al. (2014).
Instructional strategies	*BSCS Science: An Inquiry Approach* is a comprehensive 3-year high school curriculum for Grades 9–11. To make the materials educative, a variety of teacher supports were integrated including practical strategies for implementing the 5E 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). Source: Taylor, Getty, Kowalski, Wilson, Carlson, and Van Scotter (1998).
Information about students' ideas	*Project-Based Inquiry Science* is a 3-year comprehensive middle school science curriculum comprised of 8–10 week units in physical, life, and earth sciences. Materials for teachers included extensive lesson directions and different types of supports intended to be educative for teachers. For example, in one of the units activities were designed to elicit students' ideas and many possible misconceptions were listed. Many questions were provided and teachers were given suggestions on how to help (e.g., "students have read that latitude lines are like rubber bands stretched around a baseball. If students are struggling…it might help to demonstrate the lines by enacting that comparison"). Source: Schneider (2013).
Subject matter support	*Toward High School Biology* is an 8-week middle school unit that connects core chemistry and biology ideas to help students build a strong conceptual foundation for their study of biology in high school and beyond. The curriculum consists of instructional materials for both students and teachers. The suite of teacher support materials helps to deepen teachers' knowledge of science content and practices related to the unit. These materials include (among other supports) *Background Content Knowledge*, which provides teachers with more advanced information on the science content, important observations students should make, and any observations teachers might emphasize/deemphasize. Source: Kruse, Howes, Carlson, Roth, and Bourdélat-Parks (2013).

5.3 | Addressing confounds in curriculum features associated with positive outcomes

The bivariate correlational analysis of curriculum features presented in Table 7 shows a moderate correlation between scope and the presence of information about students' ideas and a weak but significant correlation between scope and the presence of subject matter support on the other hand. To explore potential confounds deriving from these correlations with respect to student outcomes, two separate logistic regression analyses were conducted (see Table 10). The first logistic regression revealed that, despite their moderate correlation, both scope and the presence of information about student's ideas make independent contributions to the prediction of student outcomes. The second logistic regression revealed that, while scope made a statistically significant contribution to the prediction of student outcomes, the effect of subject matter support was no longer significant. Although moderate correlations were also found between subject matter support and other key curriculum features (i.e., recommended instructional strategies,

TABLE 10 Exp(b) (i.e., odds ratio) from logistic regression analyses of the effects of key curriculum features on learner outcomes

	Included predictors	Exp(b)
Model 1	Scope	7.25*
	Information about students' ideas	4.39*
Model 2	Scope	9.26*
	Subject matter support	1.97

*$p \leq .05$, one-tailed.

alignment to standards, and information about students' ideas), there was no need to run further multiple regressions as the results of this second logistic regression ruled out subject matter support as a key predictor.

6 | DISCUSSION

Around the world, large investments are made in curriculum materials with the goal of supporting science education reform. However, relatively little evidence is available about what features of curriculum materials really matter to impact student and teacher learning. To address this need, the current study sought to examine curriculum features associated with student and teacher outcomes across a large sample of research-based K–12 science curriculum materials. Below, we first reflect on the key findings of the study in light of recent literature, and then discuss implications for curriculum developers, science education researchers, and those educational professionals who select or supplement curriculum materials.

6.1 | Teacher supports matter for both student and teacher outcomes

The current study reveals that curriculum materials which included recommended instructional strategies were more likely to yield positive teacher and student outcomes. This finding provides further support for previous work related to the impact of teacher supports embedded in curriculum materials. First, it adds to existing evidence about the opportunities that teacher supports provide to expand teachers' PCK and their repertoire of instructional practices (Grossman & Thompson, 2008; Schneider, 2009; Schneider & Krajcik, 2002). Second, it corroborates that these opportunities for teacher learning may, in turn, have a positive impact on student learning (Arias et al., 2017; Bismack et al., 2015; Cervetti et al., 2015). The current study also demonstrates a moderate association between student outcomes and information about students' ideas (e.g., prerequisite knowledge or alternative conceptions). Together, these findings confirm the important role of teacher supports in promoting reforms in science education.

Further, our study reveals a moderate association between teacher outcomes and the availability of information about the national or state standards addressed in the curriculum materials. Such information can support teachers in establishing connections between the ideas addressed in different units and how these build on one another (Carlson et al., 2014; Davis & Krajcik, 2005). Hence, standards information may contribute to the coherence of enacted curricula.

6.2 | The scope of curriculum materials has an impact on student outcomes

The current study also reveals a moderate association between the scope of curriculum materials and student outcomes. Specifically, curriculum materials with a larger scope (e.g., comprehensive curricula, curriculum sequences) were more likely to yield positive student outcomes. This finding could be explained in terms of curriculum coherence, defined as the adequate coordination of content and scientific practices *within* and *across* curriculum units and years of instruction (Fortus & Krajcik, 2012). While larger scope does not guarantee increased coherence, previous studies suggest that curriculum materials consisting of sequenced, coordinated curriculum units can make connections between interrelated ideas explicit, thereby helping students develop deeper and broader understandings (Fortus et al., 2015; Stevens et al., 2015). Making such connections explicit becomes particularly important to realize the vision of science learning underlying recent science reforms, but also increasingly challenging in an age of open access, digital resources, and school-based curriculum development. Although difficult to consistently assess in the available sample materials with large scope included in our sample, curriculum coherence appears to have been an explicit goal in the majority of them. We add the caution that a curriculum of large scope that is incoherent is unlikely to improve student outcomes.

6.3 | Curriculum features not associated with positive outcomes

In contrast to our initial hypotheses based on prior research, our results revealed no significant associations between student outcomes and the presence of cognitive or accommodation supports. Similarly, no associations were found between teacher outcomes and the scope of curriculum materials, or between teacher outcomes and the presence of educative supports such as information about students' ideas or subject matter support.

One cannot draw strong conclusions based on the absence of effects, since there are always concerns about statistical power (Quertemont, 2011). Nevertheless, in inspiring future studies, we wonder to what extent the absence of positive associations between supports for student learning and student outcomes may be related to the quality of the supports and how they are embedded in the curriculum materials. For example, it may be the case that the accommodation supports were poorly designed, on average. In addition, sample differences between the prior literature and the samples used in the evaluation studies might also explain differences in outcomes. For example, based on prior research, it is likely that there are heterogeneous effects for different forms of supports for student learning and on different student populations (cf. McNeill & Krajcik, 2009; McNeill et al., 2006; Reisslein, Sullivan, & Reisslein, 2007).

Regardless of the causes, the presence and absence of effects in our analyses can be taken as important suggestions for designers as to which features are most robustly associated (across possible variation in populations and implementation quality) with positive student outcomes.

6.4 | Implications

The results of the current study have several implications for curriculum developers, educational professionals who select or supplement curriculum materials, and science education researchers. First, curriculum developers should more consistently include educative supports for expanding teachers' PCK for students' ideas (e.g., information about prerequisite knowledge, alternative conceptions), which was shown here to have significant benefits for student learning. Developers may also harness new possibilities offered by online or digital materials to provide the kinds of supports that are associated with teacher learning including demonstration of instructional strategies (e.g., videos, simulations, chats with experts) or alignment to standards (easily accessible and explored through teacher friendly databases). New technologies can also enable designers to adapt the supports to the preferences and evolving needs of teachers (Krajcik & Delen, 2017).

Our findings on scope may be more complex for designers to implement given the movement toward open-source and digitally distributed curriculum materials that can be used in a very piecemeal fashion. Here, curriculum developers need to go further than making the goals of each curriculum unit explicit, by describing how different curriculum units may relate to and build on each other to support the development of science and engineering practices, disciplinary core ideas, and crosscutting concepts across one or several years of instruction. Examples of different coherent sequences and explicit descriptions of curriculum prerequisites may also be useful in this regard.

Second, educational professionals who select or supplement curriculum materials should make sure that the specific types of teacher supports shown here to have a significant effect on student and teacher outcomes are included. If not, a possible approach could be to provide professional development opportunities that may help to compensate for weakness and gaps in support within materials already selected. However, if these are not fully addressed prior to, or immediately accessible at the time of need, it does not seem likely that professional development will compensate sufficiently to yield the desired effects on student and teacher outcomes.

Finally, science education researchers should pay increased attention to the investigation of the effects of curriculum materials on diverse teacher outcomes. Our review of project documentation revealed that relatively few studies investigated issues related to the impact of curriculum materials on teacher beliefs, PCK and instructional practices. Given the interactive relationship between teachers and curriculum materials (Remillard, 2005), such studies are important to understand the ultimate effects of curriculum materials. Further, while this study sheds light on critical curriculum features that impact teacher and student outcomes, there are other dimensions along which the impact of curriculum materials could be measured. For example, a number of researchers have argued for the investigation of the scalability (rather than just student impact) of curriculum innovations, including dimensions such as sustainability, spread, and shift in ownership (Clarke & Dede, 2009; Coburn, 2003). However, our extensive review of project documentation revealed that it was difficult to find any evidence about any of these other dimensions, even when including the possibility of relatively indirect evidence.

6.5 | Closing remarks

The present study offers a modest but important contribution to the existing literature on science curriculum materials by identifying critical curriculum features that are associated with positive student and teacher outcomes. Specifically, our findings reveal that teacher supports, rather than student supports, had positive impacts on both student and teacher outcomes, and that materials with a larger scope had positive impacts on student outcomes. These results offer valuable information for future curriculum research and development efforts that meet the ambitious vision of science learning set by recent science education reform movements around the world.

ACKNOWLEDGMENTS

The research reported here was supported by the National Science Foundation, through Grants #1251562, 1252416, and 1252373 to the University of Pittsburgh, TERC, and the Lawrence Hall of Science at the University of California-Berkeley. The opinions expressed are those of the authors and do not represent views of the National Science Foundation. The authors wish to thank Jacqueline Barber, Debra Bernstein, Anushree Bopardikar, Brian Drayton and Sara Walkup for their constructive input during the execution of this study.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Natalie Pareja Roblin http://orcid.org/0000-0001-9809-9931
Christian Schunn http://orcid.org/0000-0003-3589-297X
Susan McKenney http://orcid.org/0000-0002-6028-4098

REFERENCES

Akerson, V., Nargund-Joshi, V., Weiland, I., Pongsanon, K., & Avsar, B. (2014). What third-grade students of differing ability levels learn about nature of science after a year of instruction. *International Journal of Science Education*, *36*(2), 244–276.

Arias, A. M., Smith, P. S., Davis, E. A., Marino, J. C., & Palincsar, A. S. (2017). Justifying predictions: connecting use of educative curriculum materials to students' engagement in science argumentation. *Journal of Science Teacher Education*, *28*(1), 11–35.

Arias, A., Davis, E., Marino, J., Kademian, S., & Palincsar, A. (2016). Teachers' use of educative curriculum materials to engage students in science practices. *International Journal of Science Education*, *38*(9), 1504–1526.

Ball, D., & Cohen, D. (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.

Belland, B., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, *48*(4), 243–270.

Belland, B., Walker, A., Olsen, M., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. *Educational Technology & Society*, *18*(1), 183–197.

Berland, L., & Reiser, B. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26–55.

Beyer, C., & Davis, E. (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., Delgado, C., Davis, E., & 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.

Bismack, A., Arias, A., Davis, E., & Palincsar, A. (2015). Examining student work for evidence of teacher uptake of educative curriculum materials. *Journal of Research in Science Teaching*, *52*(6), 816–846.

Brown, M. (2009). The teacher–tool relationship. Theorizing the design and use of curriculum materials. In J. Remillard, B. Herbel-Eisenmann, & G. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17–36). New York, NY: Routledge.

Carlson, J., & Anderson, R. (2002). Changing teachers' practice: Curriculum materials and science education reform in the USA. *Studies in Science Education*, *37*(1), 107–135.

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

Cervetti, G., Barber, J., Dorph, R., Pearson, P., & Goldschmidt, P. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching*, *49*(5), 631–658.

Cervetti, G., Kulikowich, J., & Bravo, M. (2015). The effects of educative curriculum materials on teachers' use of instructional strategies for English language learners in science and on student learning. *Contemporary Educational Psychology*, *40*, 86–98.

Clark, D., Touchman, S., Martinez-Garza, M., Ramirez-Marin, F., & Drews, T. (2012). Bilingual language supports in online science inquiry environments. *Computers & Education*, *58*, 1207–1224.

Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the River City curriculum. *Journal of Science Education and Technology*, *18*(4), 353–365.

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12.

Cohen, D., & Ball, D. (1999). *Instruction, capacity and improvement* (CPRE Research Report Series RR-43). Philadelphia, PA: CPRE Publications.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Dias, M., Eick, C., & Brantley-Dias, L. (2011). Practicing what we teach: A self-study in implementing an inquiry-based curriculum in a middle grades classroom. *Journal of Science Teacher Education*, *22*(1), 53–78.

Donna, J. D., & Hick, S. R. (2017). Developing elementary preservice teacher subject matter knowledge through the use of educative science curriculum materials. *Journal of Science Teacher Education*, *28*(1), 92–110.

Ellins, K., Ledley, T., Haddad, N., McNeal, K., Gold, A., Lynds, S., & Libarkin, J. (2014). EarthLabs: Supporting teacher professional development to facilitate effective teaching of climate science. *Journal of Geoscience Education*, *62*(3), 330–342.

Fogleman, J., McNeill, K. L., & Krajcik, J. (2011). Examining the effect of teachers' adaptations of a middle school science inquiry-oriented curriculum unit on student learning. *Journal of Research in Science Teaching*, *48*(2), 149–169.

Fortus, D., & Krajcik, J. (2012). Curriculum coherence and learning progressions. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 783–798). Dordrecht, The Netherlands: Springer.

Fortus, D., Sutherland Adams, L., Krajcik, J., & Reiser, B. (2015). Assessing the role of curriculum coherence in student learning about energy. *Journal of Research in Science Teaching*, *52*(10), 1408–1425.

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

Gaydos, M. J., & Squire, K. D. (2012). Role playing games for scientific citizenship. *Cultural Studies of Science Education*, *7*(4), 821–844.

Grossman, P., & Thompson, C. (2008). Learning from curriculum materials: Scaffolds for new teachers? *Teaching and Teacher Education*, *24*(8), 2014–2026.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391.

Harris, C., Penuel, W., D'Angelo, C., DeBarger, A., Gallagher, L., Kennedy, C., … Krajcik, J. (2015). Impact of project-based curriculum materials on student learning in science: Results of a randomized controlled trial. *Journal of Research in Science Teaching*, *52*(10), 1362–1385.

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, *39*(9), 870–888.

Hayes, K., & Trexler, C. (2016). Testing predictors of instructional practice in elementary science education: The significant role of accountability. *Science Education*, *100*(2), 266–289.

Janssen, F., Westbroek, H., Doyle, W., & Van Driel, J. (2013). How to make innovations practical. *Teachers College Record*, *115*(7), 1–43.

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.

Knight, V., Spooner, F., Browder, D., Smith, B., & Wood, C. L. (2013). Using systematic instruction and graphic organizers to teach science concepts to students with autism spectrum disorders and intellectual disability. *Focus on Autism and Other Developmental Disabilities*, *28*(2), 115–126.

Kowalski, S. M., Van Scotter, P., Stuhlsatz, M. A., & Taylor, J. A. (2010). Instructional materials, equity, and science achievement, Paper presented at the 2010 meeting of the American Educational Research Association, Denver, CO.

Krajcik, J., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, *25*(2), 157–175.

Krajcik, J., & Delen, I. (2017). The benefits and limitations of educative curriculum materials. *Journal of Science Teacher Education*, *28*(1), 1–10.

Kruse, R., Howes, E. V., Carlson, J., Roth, K., & Bourdélat-Parks, B. (2013). Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: Designing professional development to support teaching, Paper presented at the NARST Annual International Conference, Rio Grande, Puerto Rico.

Lachapelle, C. P., Cunningham, C. M., Jocz, J., Kay, A. E., Phadnis, P., & Sullivan, S. (2011). Engineering is elementary: An evaluation of year 6 field testing. Paper presented at the NARST Annual International Conference, Orlando, FL (Vol. *2011*).

Lederman, N. G., Antink, A., & Bartos, S. (2014). Nature of science, scientific inquiry, and socio-scientific issues arising from genetics: A pathway to developing a scientifically literate citizenry. *Science & Education*, *23*(2), 285–302.

Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W., & Lobene, E. V. (2014). Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences*, *264*, 4–18.

Marco-Bujosa, L., McNeill, K., González-Howard, M., & Loper, S. (2017). An exploration of teacher learning from an educative reform-oriented science curriculum: Case studies of teacher curriculum use. *Journal of Research in Science Teaching*, *54*(2), 141–168.

Marx, R., Blumenfeld, P., Krajcik, J., Fishman, B., Soloway, E., Geier, R., & Tal, R. (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.

McNeill, K., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Sciences*, *18*(3), 416–460.

McNeill, K., Lizotte, D., Krajcik, J., & Marx, R. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, *15*(2), 153–191.

Meyer, X., & Crawford, B. A. (2011). Teaching science as a cultural way of knowing: Merging authentic inquiry, nature of science, and multicultural strategies. *Cultural Studies of Science Education*, *6*(3), 525–547.

Meyer, X., & Crawford, B. (2015). Multicultural inquiry toward demystifying scientific culture and learning science. *Science Education*, *99*(4), 617–637.

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

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

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*(3), 241–286.

Park, S., Jang, J. Y., Chen, Y. C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, *41*(2), 245–260.

Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, *13*(3), 423–451.

Quertemont, E. (2011). How to statistically show the absence of an effect. *Psychologica Belgica*, *51*(2), pp. 109–127.

Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, *13*(3), 273–304.

Reisslein, J., Sullivan, H., & Reisslein, M. (2007). Learner achievement and attitudes under different paces of transitioning to independent problem solving. *Journal of Engineering Education*, *96*(1), 45–56.

Remillard, J. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211–246.

Remillard, J., Harris, B., & Agodini, R. (2014). The influence of curriculum material design on opportunities for student learning. *ZDM Mathematics Education*, *46*(5), 735–749.

Rivet, A., & Krajcik, J. (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.

Roseman, J. E., Fortus, D., Krajcik, J., & Reiser, B. J. (2015). Curriculum materials for Next Generation Science Standards: What the science education research community can do. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, Chicago, IL.

Sadler, T., Romine, W., Menon, D., Ferdig, R., & Annetta, L. (2015). Learning biology through innovative curricula: A comparison of game-and nongame-based approaches. *Science Education*, *99*(4), 696–720.

Schneider, R. (2013). Opportunities for teacher learning during enactment of inquiry science curriculum materials: Exploring the potential for teacher educative materials. *Journal of Science Teacher Education*, *24*(2), 323–346.

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

Schneider, R., 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.

Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Achér, A., Fortus, D., … Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654.

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

Sinha, S., Gray, S., Hmelo-Silver, C., Jordan, R., Eberbach, C., Goel, A., & Rugaber, S. (2013). Conceptual representations for transfer: A case study tracing back and looking forward. *Frontline Learning Research*, *1*(1), 3–23.

Skaalvik, E., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education*, *26*(4), 1059–1069.

Stevens, S., Delgado, C., & Krajcik, J. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, *47*(6), 687–715.

Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (2015). An efficacy trial of research-based curriculum materials with curriculum-based professional development. *American Educational Research Journal*, *52*(5), 984–1017.

White, B., & Frederiksen, J. (1998). Inquiry, modelling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118.

White, B., & Frederiksen, J. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. In J. Minstrell & E. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 331–370). Washington, DC: American Association for the Advancement of Science.

Wyner, Y. (2013). The impact of a novel curriculum on secondary biology teachers' dispositions toward using authentic data and media in their human impact and ecology lessons. *Journal of Science Teacher Education*, *24*(5), 833–857.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Roblin PN, Schunn C, McKenney S. What are critical features of science curriculum materials that impact student and teacher outcomes?. *Sci Ed*. 2018;102:260–282. https://doi.org/10.1002/sce.21238