NGSS-Designed Curriculum Materials for the Middle Grades: Results from a Study on Student Learning in Physical Science

Robert F. Murphy (iD)

LFC Research

Christopher J. Harris

WestEd

Mingyu Feng

WestEd

Ashley Iveland

WestEd

Melissa Rego

WestEd

Daisy Rutstein

edCount, LLC

Kevin Huang (iD)

WestEd

A new generation of science curriculum materials is being developed and made widely available across K–12. These materials, along with their associated professional learning, are expected to play a pivotal role in transforming science instruction to meet the vision of A Framework for K–12 Science Education and the performance expectations of the Next Generation Science Standards. This study investigated one of the first widely available middle school curriculum programs designed for the NGSS, Amplify Science 6–8. The study team conducted a randomized experiment to examine the implementation and efficacy of the middle school curriculum program (materials plus teacher professional learning) for improving seventh grade students' learning in physical science. Findings show that students in schools where the program was implemented significantly outperformed peers in the comparison condition on an NGSS-focused physical science assessment. These findings underscore the importance of conducting additional research on NGSS-designed curriculum programs across other grade levels and within different science domains.

Keywords: *Experimental design, experimental research, hierarchical linear modeling, instructional practices, middle schools, mixed methods, multi-site studies, Next Generation Science Standards, program evaluation, professional development, randomized controlled trial, science curriculum, science education*

Released over a decade ago, a Framework for K–12 Science Education (Framework) (National Research Council [NRC], 2012) and the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) set forth a vision and foundation for K–12 science education that is considerably different from previous conceptualizations of science learning. In the Framework and NGSS, the science education community emphasizes a view of learning as a process of using and

applying disciplinary core ideas (DCIs) in concert with science and engineering practices (SEPs) and crosscutting concepts (CCCs) to make sense of phenomena or to solve problems. Central to this vision is the notion of three-dimensional learning, in which students use the three dimensions of DCIs, CCCs, and SEPs as the means through which to build the proficiencies required to meet the performance expectations of the NGSS. The performance expectations express the integrated goals for three-dimensional learning. They specify what students should know and be able to do in science at a given grade level or across a grade band.

Because the Framework and NGSS are so different from prior standards, it has taken time to develop and make widely available the curriculum materials, teacher professional learning, and assessment resources needed to advance the vision (Pellegrino et al., 2014; Penuel & Reiser, 2018). Also influencing availability has been the gradual shift by states toward adopting standards based on the Framework and NGSS. We are now at a point where the vision for science education has become part of education policy in many corners of the U.S. Nearly all states now have standards influenced by the Framework alone or both the Framework and NGSS (National Science Teaching Association, 2022). Noteworthy is that performance expectations are articulated in the standards of all states that fully adopted the NGSS and have been adapted in the science standards of many states whose standards are based on the Framework.

Increasingly, new curriculum materials are becoming available to districts and schools to support teachers in providing instructional experiences that will engage their students in three-dimensional learning. Many are being designed to meet the ambitious call of the NGSS and to address the performance expectations that are found in state science standards. As these NGSS-designed curricula are being implemented more widely, it is important to conduct research on their efficacy to (1) gather evidence as to whether these new curriculum materials accompanied with professional learning can enhance student learning of multi-dimensional learning goals found in state science standards and (2) to inform state-, district-, and school-level decision-making about how best to support the implementation of instruction that meets today's vision for science education.

This article describes findings from a study of a middle school science curricular program that was designed to promote learning as called for by the Framework and NGSS. The widely available materials, Amplify Science 6–8 (AS), were developed by the University of California, Berkeley's Lawrence Hall of Science in collaboration with Amplify Education, Inc. AS materials are among the first comprehensive curricular programs designed specifically to meet the vision of the Framework and to address the performance expectations of the NGSS. The materials received high marks for their NGSS design in an independent review by EdReports (2020). Also noteworthy is that the materials have reached broad scale; they can be found in use in every U.S. state.[1]

The study team set out to investigate the extent to which the AS curriculum program enhances students' three-dimensional learning and teachers' instruction to support this type of learning. We conducted a randomized controlled trial in seventh grade science classrooms within three public school districts across two states. The research reported in this article focuses on understanding the impact of the use of the AS curriculum program on teachers' instructional practice and students' science learning in physical science.

Background

Today's Vision for Science Education

Over the past decade, the Framework and NGSS have transformed science education policy and practice in schools throughout the country (Anderson et al., 2018). The Framework and NGSS specify what students should know as well as what students should be able to do with what they know in science at a given grade level or across a grade band. A main tenet is that science proficiency develops over time and becomes more robust when students have opportunities to put knowledge into use. This is quite different from the more familiar format of science learning where building proficiency is considered primarily a matter of acquiring knowledge. Within the realm of today's vision for science education, it is not solely what students know but also how they use and apply what they know that helps to advance learning.

The Framework describes three interconnected dimensions of science proficiency—DCIs, CCCs, and SEPs. DCIs denote the big comprehensive ideas that are associated with a discipline, like matter and its interactions in physical science, and that are essential to explaining a wide range of phenomena (Duncan et al., 2017). CCCs are ideas such as systems thinking and cause and effect that are important across many science disciplines and provide a unique lens to examine phenomena (Nordine & Lee, 2021). SEPs are the multiple ways of knowing and doing—for example, developing models, analyzing and interpreting data, and constructing explanations—that scientists and engineers use to study the natural and designed world (Schwarz et al., 2017). The SEPs require that students do the "walk and talk" of science and engineering in ways similar to what is done in professional practice but in ways appropriate for students. The Framework focuses on the need for the integration of these three dimensions in science and engineering education. As students engage with the dimensions, they use their knowledge in varied and demanding ways and through this process they transform that knowledge. Accordingly, all students must have the opportunity to learn and actively participate in science through using the three dimensions (NRC, 2014). Moreover, the integrated use of

the dimensions should be integral to assessing what students know and can do (NRC, 2014, 2015).

A central position of the Framework is that proficiency is demonstrated through performances that require the integration of all three dimensions. Such performances are referred to as performance expectations and they are articulated as a set of standards in the NGSS. The performance expectations incorporate specific actions like "analyze and interpret," "develop and use a model," "design a solution," or "construct an explanation," in which the practices of science and engineering are integrated with DCIs and CCCs. An example performance expectation from the middle school grade band for physical science is: MS-PS1-2. Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred (NGSS Lead States, 2013). This performance expectation requires that students grasp a number of important chemistry ideas (e.g., substances and their properties, how substances can interact, among others) and be adept in making sense of data including analyzing data to identify relevant patterns that relate to their knowledge about chemical reactions.

Realizing the vision of the Framework is an ongoing effort that has required major changes in state standards and state large-scale assessments; changes in district-level policies, practices, and benchmark assessments; new formats for teacher professional learning; and changes in classroom-based resources for teachers and students. Among the most vital supports for teachers and students to bring the vision for science education to their classrooms are curriculum materials. There are many prior studies that provide evidence that curriculum materials matter for teachers' instructional practice and for students' learning in science (e.g., Harris et al., 2015; Huffman et al., 2003; Marx et al., 2004; Taylor et al., 2015; Wilson et al., 2010). However, there have been few studies to date on the impact of curriculum materials designed to support the ambitious teaching and learning required of the Framework and NGSS.

Evidence of the Role of Curriculum Materials in Supporting the Vision

In today's science classrooms, an important role of the teacher is to activate, monitor, and support students' three-dimensional learning by providing instructional experiences that will engage their students in using and applying the three dimensions of science proficiency. Prior research highlights the impactful role that curriculum materials can play in supporting teachers and students in making shifts in classroom practice. Well-designed science curriculum materials provide important resources for teachers including routines, instructional strategies, and discussion prompts that can help them take up new formats for instruction (e.g., Harris et al.,

2012; McNeill, 2009; Roblin et al., 2018) and provide opportunities for them to learn themselves as they teach (Davis & Krajcik, 2005; Krajcik & Delen, 2017). Research has also shown that curricular materials and professional learning together can enhance teachers' ability to engage students in ambitious science learning (e.g., Lee et al., 2008; Penuel et al., 2011; Short & Hirsh, 2020; Taylor et al., 2015). For students, curriculum materials are widely acknowledged in the research literature for their central role in supporting learning (Geier et al., 2008; Harris et al., 2015; Taylor et al., 2015).

Contemporary materials designed to support new modes of learning, such as the three-dimensional learning called for by the NGSS, increasingly include structures to engage students in activities in ways similar to how scientists and engineers conduct their work along with embedded scaffolds for doing so and with supported practice in reading, writing, and speaking the discourses of science (Penuel & Reiser, 2018). In recent years, several comprehensive curricula have been expressly created to align with the vision of the Framework and NGSS. These curricula are accompanied with professional learning support for teachers so they can learn how to use the materials in ways that cohere with the stance of three-dimensional learning (e.g., Short & Hirsh, 2020). The new generation of curriculum materials are just now becoming widely available and many have not been tested at scale using rigorous study designs. Prior to 2020, very few NGSS-aligned curriculum materials existed. Perhaps not surprising, until now school districts have been using a range of curriculum materials with varying alignment to the vision (Lowell et al., 2021), including inquiry-based materials that were developed before the release of the NGSS and then updated or "redesigned" for the NGSS, materials that districts developed themselves as part of internal efforts to produce their own homegrown resources, textbooks that have been reconfigured for the NGSS, and open-source and fee-for-service online programs developed to align with NGSS (some of these developed with funding from federal grants and foundations). Smith (2020), for example, reported data from the 2018 National Survey of Science and Mathematics Education indicating that many science teachers by then were still using materials published before the release of the Framework and NGSS. Noteworthy too is that teachers have also attempted to adapt their outdated curricular units and lessons to meet the requirements of the vision, with varying degrees of success.

A challenge for the science education field has been that few ready-for-scale, NGSS-aligned materials have been available for a long enough period to achieve broader use and thus warrant an investment in efficacy research. Some earlier efficacy studies, such as one conducted by Harris and colleagues (Harris et al., 2015), examined the effectiveness of middle school science curriculum materials with teacher professional learning that had some features that matched

well with the vision for science learning put forth by the Framework. More recently, Krajcik et al. (2023) studied upper elementary science curriculum materials that they had developed and that were designed expressly to support three-dimensional learning and achievement of performance expectations. However, the number of these types of studies has remained low. The encouraging news is that this is now changing—school districts across the country are adopting curriculum materials that have been designed with the Framework and NGSS as their foundation. As these materials achieve scale, efficacy studies, like the one that is the focus of the article, are needed to develop estimates of their impacts. Importantly, these studies need to measure learning outcomes using assessments that elicit integrated performance of the SEPs, DCIs, and CCCs (DeBarger et al., 2016; NRC, 2014).

Methods

In this study, we investigated the comprehensive middle grades science curriculum, Amplify Science 6–8, that was designed to support teachers and students in achieving the NGSS's vision for science learning. We employed an efficacy study design where we sought to examine the curriculum program (i.e., the curriculum materials plus related professional learning) under ideal conditions (Institute of Education Sciences & National Science Foundation, 2013) to determine the impacts on student science learning as well as gain insight into teacher uptake of the materials. An efficacy study of the curriculum was warranted given that the curriculum is fully developed and widely distributed and used, but has not as of yet been independently assessed on whether and under what conditions it has the intended impacts on student learning.

Research Questions

The efficacy study was guided by the following research questions examining student learning and curriculum implementation:

1. *Student Learning*:
   (RQ 1a). What is the impact of the curriculum program (i.e., AS materials and associated teacher professional learning) on NGSS-focused learning outcomes?
   (RQ 1b). How does the impact vary by student background characteristics?
2. *Curriculum Implementation*:
   (RQ 2a). What is the nature of teachers' implementation of the AS curriculum?
   (RQ 2b). In what ways does uptake of the curriculum program influence teachers' NGSS instruction?

Research Design

The research team implemented a multisite cluster-randomized controlled trial design to test the efficacy of the use of AS curriculum materials and associated professional learning. At the outset, twenty-nine middle schools were recruited across three districts for the experimental study. The districts were located in two states, one in the West and other in the Midwest. Both states had adopted the NGSS performance expectations as their middle school science standards. The three districts were committed to implementing the NGSS in 7th grade science instruction during the 2019–2020 academic year. The study team worked closely with the districts to communicate to school leaders and teachers about the study and participant obligations. Prior to the randomization of schools, school leaders and teachers interested in participating were asked to sign a participant agreement that outlined the benefits, incentives, and obligations of study participation as well as notifying them that schools would be randomly assigned to the treatment or control condition. Only schools where principals and teachers signed the agreement were included in the study. Schools were randomly assigned within-district to an immediate use of the AS curriculum condition (treatment) or a business-as-usual condition (control). Prior to random assignment, schools within each district were matched on school total enrollment, percent of black and Hispanic students, percent of students who are socioeconomically disadvantaged, and the average percent of students meeting or exceeding state standards in ELA, math, and science. Within each district, we used an optimal Mahalanobis matching approach to pair schools based on the matching variables (Gu & Rosenbaum, 1993; Rubin & Thomas, 2000).[2] Schools were grouped into matched pairs or a triad and then randomly assigned to condition (15 to treatment and 14 to control).[3] Schools remained in their assigned condition for one full academic year. As an incentive for participation in the study, the control schools were offered access to the AS materials and teacher professional learning to support its use after the main study concluded.

Analytic Sample

The study was underway during the 2019–2020 school year when the COVID-19 pandemic caused school closures in early spring 2020. This study reports the findings from 18 schools (the "analytic" sample) where teachers were able to complete their lessons in physical science and administer the study's post assessment prior to school closures due to the onset of the pandemic (10 treatment schools and 8 control schools). Fifteen of the schools were located within two districts in two large cities with populations of greater than 250,000. The remaining 3 schools were located within a single district in a small suburban city with a population of

TABLE 1
School demographics for the 18 schools in the analytic sample

Variables	All		Treatment		Control	
	Mean	Range	Mean	Range	Mean	Range
Size (7th grade enrollment)	236.8	84–514	208.5	80–305	272.1	84–514
Socioeconomic disadvantage[a]	43.6%	11.4–92.8%	43.0%	18.0–64%	44.2%	11.0–93.0%
English language learner status	7.5%	0.0–28.3%	5.6%	0.0–11.2%	10.0%	0.0–28.3%
Asian	8.9%	1.7–28.1%	7.6%	1.23–20.6%	10.5%	1.72–28.1%
Black	5.5%	1.1–19.6%	5.9%	1.1–19.6%	5.0%	1.2–13.5%
Hispanic	25.5%	12.3–76.3%	24.4%	8.9–43.8%	26.9%	9.1–76.3%
White	41.6%	1.3–78.2%	48.0%	11.4–78.2%	33.7%	1.3–76.8%
Other	9.5%	0.0–62.4%	4.5%	0.0–25.7%	15.7%	0.0–62.4%
Grade 7 Math[b]	55.1%	29.0–75.4%	54.2%	20.0–75.0%	56.2%	20.0–75.0%
Grade 7 ELA[b]	60.4%	31.0–81.0%	59.8%	43.0–74.4%	61.2%	31.0–81.0%

[a]Percent students eligible for free or reduced-price lunch.
[b]Percent students meeting or exceeding state standards in 7th grade, based on data from the 2016–2017 or 2017–2018 school year.

less than 100,000. A majority of the schools in the analytic sample (11 of 18) qualified for schoolwide Title I funding. (i.e., 40% or greater of the students in these 11 schools qualified for free or reduced-price lunches). Table 1 shows the school-level demographics for the 18 schools in the analytic sample. The sample included 19 science teachers within the 10 schools assigned to the treatment condition and 14 science teachers within eight schools assigned to the control condition. Within the participating classrooms, 1,953 students completed a physical science assessment.

Amplify Middle School Curriculum and Professional Learning (Treatment Condition)

Teachers in the intervention group implemented the AS curriculum and received professional learning provided by the Lawrence Hall of Science curriculum developers. Treatment teachers were expected to use the AS materials as their primary curriculum materials. The curriculum is designed around four primary design goals and a set of features to support those goals. The four goals and supporting design features, described in Table 2, emphasize learning by developing complex causal explanations of phenomena; strong support for science and literacy integration via reading, writing, and discourse activities; the use of technology including digital simulations; and placing students into the roles of scientists.

The program includes a digital platform for teachers and students along with physical materials for hands-on activities. The platform for teachers includes digital instructional guides with unit descriptions and lesson plans, online monitoring and reporting tools for following student progress, and resources for facilitating instruction including teaching strategies and planning and preparation steps for hands-on activities. The student platform takes the form of a digital

workspace where students can house their work and have access to grade-appropriate science articles, science simulations, and modeling tools. The classroom-based materials include hands-on kits for investigations and print materials for every unit. Students use investigation notebooks (print and digital versions are available) to record data, reflect on ideas from texts and investigations, and construct explanations and arguments among other activities. Teachers are provided with all materials for a full year of instruction accompanied with a professional learning experience to equip them for implementation.

The teacher professional learning sessions were in-person and held at three points during the school year for a total of 24 hours. The first sessions were held over two full days prior to the opening of the school year. These sessions included a focus on the AS pedagogical approaches as well as practical orientation to the digital platform and physical materials. Teachers experienced, as a student, key routines from the first set of curricular units they would teach in the fall, and reflected on how figuring out phenomena like a scientist leads to 3-D learning. The next two sessions consisted of two half-day events held during the school year and focused on preparing teachers for the second and third (of three) unit sets. These sessions include time to reflect on the teachers' experience teaching the previous units and also to delve more deeply into aspects of the curriculum such as the explanation build and opportunities for differentiation. All treatment teachers attended the summer professional learning session and all but one treatment teacher attended the other two sessions which were held during the school year.

The AS units we studied were in physical science and addressed the comprehensive DCI, Matter and Its Interactions. Specifically, the units focus on the DCI components of structure and properties of matter including phase change, substances, and properties of substances, and

TABLE 2

Design goals and features for Amplify Science curriculum materials

---

Design Goal 1. Support students in figuring out complex causal explanations of phenomena underlying real world problems.

Design Feature 1a. **Real world problem context.** Each unit includes a real-world anchor phenomenon and a purpose for investigating and explaining that phenomenon. For example, in the Grade 7 unit Phase Change students explain the mysterious phenomenon of a disappearing methane lake on Titan, a moon of Saturn.

Design Feature 1b. **Explanation build.** Each unit is organized around a unit-specific learning progression describing levels in an increasingly complex and integrated explanation of the anchor phenomenon.

Design Goal 2. Empower all students to see themselves as scientists.

Design Feature 2a. **Explicit student role.** Students take on a role as a science and engineering professional, working collaboratively to address the unit's central problem. For example, in the Grade 7 unit *Phase Change*, students' role is chemists working for a space agency.

Design Feature 2b. **Diverse role models.** Materials (books, articles, videos) explicitly attempt to overrepresent groups that have traditionally been underrepresented in science and engineering, including people of color, women, LGBTQ+ individuals, and people with disabilities.

Design Goal 3. Provide access to deep learning for all students.

Design Feature 3a. **Explicit support for engagement in the disciplinary literacy of science.** Materials provide explicit support to both students and teachers for the literacy and language demands of engaging in science practices. Routines and scaffolds include active use of a limited, core set of key vocabulary; provision of multi-language glossaries; activities that expose multiple-meaning words; and mentor texts that model scientific language use.

Design Feature 3b. **Multiple opportunities to engage with key ideas using multimodal evidence sources.** For each key idea, students encounter evidence and express their understanding in multiple modalities using the Do-Talk-Read-Write-Visualize approach adapted from Pearson and colleagues (Pearson et al., 2010), including first hand investigations; secondary evidence from texts; and student-to-student discourse.

Design Goal 4. Enable investigation of phenomena beyond what is typically possible in the classroom.

Design Feature 4a. **Custom digital simulations co-designed with other curriculum materials and activities.** Custom interactive simulations were designed for each unit providing students with closer access to underlying processes and mechanisms otherwise inaccessible in the classroom. For example, in the Grade 7 *Chemical Reactions* simulation, students can observe properties of many substances, combine substances and observe the results, and vary the quantity of substances combined. The design of the simulations was intertwined with the design of the curriculum, leveraging synergies between the design of the simulations themselves, the curricular activities in which they are used, and the other activities in the unit.

Source, Loper et al. (2022).

chemical reactions including conservation of matter and role of energy. These units engage students in using and applying the three dimensions to investigate and explain anchor phenomena. For instance, in one curricular unit students investigate the anchor phenomenon of an unknown substance discovered in a community's water supply. Each unit culminates with students constructing a causal explanation of the anchor phenomenon.

Science Curriculum and Professional Learning in the Control Schools (Control Condition)

Teachers in the control condition were asked to teach with their regular curriculum materials and participate in science professional learning sessions as they typically would in their districts. The range of enacted curriculum materials varied across control schools, but all were focused on NGSS instruction. Teachers in one district used a widely available redesigned curriculum for the NGSS; teachers in another used their own district-developed curriculum to address the NGSS performance expectations; and most of the teachers in the third district used a district-adopted textbook while some used an open-source, project-based NGSS curriculum. Curricula that are "redesigned" are materials that had been in wide use before the release of the Framework and NGSS and were subsequently updated to align with the Framework and NGSS. District-developed curricular materials are those created within school districts by science educators to meet the requirements of the Framework and NGSS. The district that used a textbook was in a transition toward a new curriculum adoption and in addition to the textbook, teachers had leeway to supplement their primary curriculum. All teachers in the comparison group were asked to implement their regular NGSS instruction on physical science topics relating to structure and properties of matter and chemical reactions. Teachers in both experimental conditions were held accountable to the same NGSS performance expectations.

Data Collection Activities, Instruments, and Measures

A series of data collection activities and instruments were deployed to collect data on the background characteristics of the participants, to understand the use of the AS materials in the treatment classrooms, to assess the contrast in instructional practices in classrooms across both experimental conditions, and to assess student learning.

Student Demographic Information. Student demographic information was collected directly from each of the three districts. This included student gender, free and reduced-price lunch status, and individualized education program (IEP) status. In addition, the evaluators collected classroom rosters from each teacher participating in the study. Rosters included information that identified the specific class, student names, and students' special education and IEP status. Prior to conducting any analyses, teacher and student names were removed from data files and replaced with anonymized research identifying numbers. Analysts worked with anonymized data files only.

Student Prior Achievement. This experiment used 7th grade students' prior 6th grade mathematics and English language arts scores on end-of-year state assessments as a measure of prior achievement. Data for statewide standardized assessments was collected directly from each district. Since the state assessments differed by state, prior to being used in the analytical models, the scores within each state were first converted to z-scores based on the respective mean score and standard deviation for the state.

Student 7th Grade Physical Science Performance. At the outset of the study there were no existing off-the-shelf assessments for the NGSS. Subsequently, the research team developed an assessment for physical science. The assessment elicits performance with aspects of NGSS performance expectations under *MS-PS1: Matter and its Interactions* and includes constructed-response tasks that address aspects of disciplinary core ideas, science and engineering practices, and crosscutting concepts. The assessment was informed by the design work of the Next Generation Science Assessment project (Harris et al., 2019) and made fair to both groups (DeBarger et al., 2016) by designing tasks to align to the performance expectations in physical science that were in the state standards of the participating schools and that all teachers were expected to teach in seventh grade. Once an initial set of tasks was developed, feedback on all tasks was obtained first through expert reviews with science education experts and second through cognitive interviews with students. On the basis of the reviews and interviews, revisions were made to the tasks to improve clarity for students and alignment to performance expectations. Finally, a pilot was conducted with 493 students to gather additional information on the performance of the tasks.

The finalized assessment consists of seven paper-and-pencil tasks that were contextualized in scenarios presented in a succinct story format with prompts to elicit integrated responses. The assessment was designed to be administered within a 50-minute class session. The SEPs addressed by the physical science tasks include developing and using models; analyzing and interpreting data; and obtaining, evaluating, and communicating information among others. Crosscutting concepts include patterns, cause and effect, and energy and matter among others. Figure 1 includes a sample item from the assessment.[4]

Teachers in both groups were requested to administer the assessment to students within two weeks after completing their instruction of the physical science topics. For scoring of the assessments, all collected assessments were randomized and assigned to independent scorers who received extensive training on the rubrics, with one set of assessments used as a training set. Scorers were blinded to students' identity and experimental condition. Aside from the training set, over 20 % of the assessments were scored by two scorers, with checks for reliability. Any disagreements were resolved by a third scorer. Scores on tasks were totaled to get an overall total score for the assessment. After completing the scoring, we examined the psychometric properties of the assessment. We found that there was a range of scores in both the treatment and control conditions, indicating that the assessment was able to capture differences in student performance. The overall internal consistency of the assessment was .788 (Cronbach's alpha). The results of a confirmatory factor analysis showed that a one-factor model provided good fit indicating it is acceptable to consider just one score from the assessment.

Curriculum Enactment and Science Instruction Context and Practices. We developed and employed administrator and teacher interview protocols, an instructional log and an end-of-year survey to investigate teachers' curriculum enactment and instruction practices in both experimental conditions and to understand the district contexts in which the science instruction was embedded.

Instructional logs. Separate instructional logs for physical science instruction were developed for treatment and control conditions. While the same core set of items used to collect information on instructional practices was included in both instruments, the instructional logs developed for treatment teachers included additional items to collect information about teachers' use of the AS curriculum materials. The instructional log included two main types of questions: (1) enactment questions that focused on the topics taught, the source of the curriculum materials, and, for teachers in the treatment condition, the specific AS

2. Miranda had 2 liquids in different test tubes. She mixed them together and then heated them to see what would happen. After heating, two separate layers formed - Layer A and Layer B.
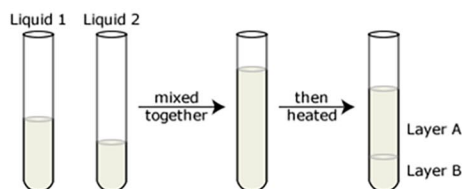
Miranda conducted tests and took measurements of the liquids and layers and recorded the data in Table 1.

Table 1. Data on substances before and after heating.

	Sample	Volume (cm³)	Solubility in Water	Boiling Point (°C)
Before Heating	Liquid 1	0.45	Yes	78.4
	Liquid 2	0.35	Yes	118.5
After Heating	Layer A	0.40	No	77.1
	Layer B	0.30	Yes	100.0

a) Based on the data in Table 1, did a chemical reaction occur?
    ☐ Yes
    ☐ No

b) **What** evidence in the table supports your choice?
c) Explain **why** the evidence you stated in Part B supports your answer in Part A. Include in your response how properties of substances help determine whether a chemical reaction occurred.

FIGURE 1. *Sample item from the student 7th grade physical science performance assessment.*
Note: This task was adapted from the Next Generation Science Assessment collaborative which was supported with funding by the National Science Foundation (Grant numbers 1316874, 1316903, 1316908 & 1903103).

lessons and activities used and any modifications made; and (2) instructional questions that focused on the frequency of engaging students with the NGSS dimensions, general scientific practices, a variety of independent and collaborative instructional opportunities as well as a set of typical challenges teachers might have encountered in teaching their science lessons in a particular week. Teachers in both groups completed the instructional logs on a weekly basis during their instruction on topics in physical science. The logs were completed online, similar to a typical online survey, and were designed to be completed in 10–15 minutes. All teachers were sent an email toward the end of each week with a link to the individual instructional log and were encouraged to complete the logs within three days of receiving the link. The average number of logs submitted by teachers in the treatment condition was 10 and the average number submitted in the control condition was 8.3. All participating teachers submitted one or more logs.

End-of-year teacher survey. The end-of-year survey was used to collect retrospective teacher self-reports of their instructional practices, resources available to support their science instruction including professional learning, various factors that may have impacted their ability to teach science, and the extent to which they engaged students in the NGSS dimensions. For the teachers in the treatment condition, we also used the survey to collect data on factors that support or hinder the implementation of AS, their satisfaction with training and follow-up support, their perceptions of the curriculum's impacts on student learning and achievement, and self-reported changes in their practice due to the intervention. The end-of-year teacher survey was administered online immediately following school closures (in early spring, 2020) and was designed to be completed within 20–30 minutes. Notification emails with a link to the survey were sent to all participating teachers. For the teachers included in the analytic sample, overall response rates were 86% (93% in the treatment condition and 77% in the control condition).

Administrator and Teacher interviews. Administrators and teachers were interviewed using a semi-structured interview protocol to better understand the district context for science instruction and the implementation of the AS

curriculum. For the interviews of district and school administrators, a semi-structured interview protocol was used to probe for specific policies, programs and the availability of professional learning opportunities that might influence how science is taught in both treatment and control schools along with the use of the AS curriculum in treatment schools. The semi-structured teacher interview protocol asked about teachers' experience using the AS curriculum, their perceptions of its benefits and areas for improvement, and the extent they perceived that the AS curriculum program improved their capacity to implement the NGSS. Teacher interviews were conducted after all physical science instruction had concluded and prior to the end of the school year. District science leads and a sample of school administrators were interviewed immediately after the school year concluded. We interviewed at least one science lead in each of the 3 districts, 4 school principals and 7 of the 14 treatment teachers. All interviews were conducted via phone, and each was scheduled to be completed within 30 minutes.

Amplify Science 6–8 Implementation Measures. Data collected through the instructional logs was used to assess teachers' implementation of the AS curriculum including the extent to which they used the AS materials and materials from other sources to teach topics in physical science. Each week during the teaching of physical science teachers were asked to report whether their instruction covered lessons from the AS physical science units (Phase Change and Chemical Reactions), the chapters within each unit, and to what extent they used the AS materials with or without supplementing these materials with materials and activities from other sources (e.g., other commercial curricula, district- or teacher-developed materials, etc.). Within each of the four chapters for each unit, teachers also were asked to report whether they taught each lesson with or without modification or had skipped the lesson altogether and why they may have modified or skipped the lesson. At the conclusion of each unit, teachers were asked to report any challenges in using the AS materials and implementing activities with their students and to rate the AS unit materials (e.g., *not as good, as good, better*, etc.) in comparison to other curriculum materials they had used in the past to teach the same topics.

Science Instruction Measures. We also used data collected from the instructional logs to document the nature of the science instruction provided by teachers in both conditions during the teaching of physical science. We were interested in using this data to examine whether differences in instruction between conditions might help explain any differences found in student learning. Specifically, we examined (1) teachers' coverage of NGSS SEPs and (2) the prevalence of particular kinds of instructional opportunities provided to students. Data on the extent to which teachers in each condition engaged students in the SEPs during the teaching of physical science units provides a sense of the extent to which students'

experiences in the treatment and control classrooms differed in the time they spent engaged in NGSS-promoted practices during science instruction. In a similar way, data on the extent to which students in each condition were exposed to a range of instructional opportunities allows us to investigate whether teachers in the treatment schools tended to use different instructional strategies to support their students' science learning compared to teachers in the control schools. Together these measures were used to give us a picture of how the intervention—Amplify Science materials plus professional learning—may have changed how science was taught in treatment classrooms.

For the measures of *Coverage of NGSS SEPs*, teachers were asked to report in the weekly instructional log the percentage of instructional time (from 1, *Not at all*, to 6, *More than 75%*) that they had students engage in each of the eight SEPs—*asking science questions; developing and using models; planning and carrying out investigations; analyzing and interpreting data; using mathematics and computational thinking; constructing explanations; engaging in arguments from evidence;* and *obtaining, evaluating and/or communicating information*. For the *Prevalence of Instructional Opportunities* measures, teachers were asked to report in the weekly instructional logs whether they provided their students with any of nine different instructional opportunities (*mark all that apply*)—*a hands-on experience that supported science learning; a demonstration of students to watch; opportunities for students to do individual seat work; opportunities for students to read science texts; opportunities for students to connect science learning to everyday experiences; opportunities for students to write in science notebooks (in paper or "digital" forms of notebooks); opportunities for students to communicate their scientific thinking to one another;* and *a quiz for students to complete*. For each item, a teacher's responses across multiple weekly logs were compiled and converted into the percentage of responses that the teacher indicated they provided their students with a particular instructional opportunity. Since teachers submitted multiple teacher logs, for analysis purposes, teachers' responses to individual items across logs were either converted into averages when the measurement scale was ordinal or percentages when the measurement scale was binary (*Yes* or *No*) by counting the number of times a teacher selected a particular response and then dividing by the total number of logs the teacher submitted.

Data Analyses

We conducted three types of analysis:

1. Descriptive analyses of the sample and sample attrition, and treatment teachers' implementation of the AS curriculum (RQ 2a).
2. An analysis of the impact of the AS curriculum program on student learning in physical science includ-

ing an analysis of potential moderators (RQ 1a and 1b).

3. An exploratory analysis of the impact of the uptake of the AS curriculum program on teacher practice and instruction (RQ 2b).

Descriptive Analyses. An analysis of the student sample and student-level attrition was conducted using the student demographic information and prior achievement measures (6th grade state reading and mathematics achievement scores) provided by the districts, classroom roster information provided by the districts, and records of students who completed the study's assessment administered by the research team. We tracked both students who joined and left the study after the start of the school year (joiners and leavers), to examine differential attrition rates between conditions, and to see whether the population that dropped out of the study differed systematically from the students who remained. A description of the extent to which teachers used the AS curriculum materials and their experience with the curriculum was based on a descriptive analysis of the treatment teachers' self-reported implementation data collected through the study's weekly instructional logs (RQ 2a).

Main Impact Analysis. For the main impact analysis, we investigated the main effect as well as possible moderators of impacts (RQ 1a and RQ 1b). To estimate the impact of the AS curriculum program on student learning outcomes, we compared the scores on the physical science assessment for students in treatment schools with the scores for students in control schools. The analysis used a two-level hierarchical linear regression model (students nested within schools) controlling for school-level and student-level characteristics. Student-level covariates used in the model included: prior 6th grade mathematics and English language arts (ELA) state achievement scores, gender, English Language Learner (ELL) status, individualized education program (IEP) status, and ethnicity (Asian, Hispanic, White, and Other). At the school-level we included the randomization block or strata as a covariate. All analyses were conducted with students' physical science scores as the dependent variable.

The two-level hierarchical regression model used to estimate the main impact of assignment to the use of the AS curriculum on student performance on the physical science assessment is shown below:

$$\text{Test}_{ijk} = \alpha_0 + \beta_1 \text{newPre}_{ijk} + \beta_2 \text{dPre}_{ijk} + \beta_3 \text{Tx}_{jk} + \Sigma \beta_l I_{ijk}$$
$$+ \Sigma \beta_{dl} \text{dI}_{ijk} + \Sigma \nu_s \text{Stratum}_k + \tau_{jk} + \varepsilon_{ijk}$$

where subscripts *i*, *j*, and *k* denote student, school, and stratum; *Test* represents student achievement score (total or item score); *newPre* represents the baseline measure with missing values coded to a constant; *dPre* is the missing indicator for *newPre*; *Tx* is a dichotomous variable indicating student enrollment in a school (or in a teacher's class) that has been assigned to treatment or control condition; *I* is a vector of other control variables for students, measured prior to exposure to the intervention (again, the missing values were coded to a constant); *dI* is a vector of missing indicators for *I*; *Stratum* represents a vector of fixed effects for *k–1* strata; τ represents a random variable for schools (clustering group), and ε is an error term for individual students. The intervention effect is represented by β_3, which captures treatment-control differences on the outcome variable after controlling for all covariates and study design factors (strata).

The main impact analysis was conducted in two steps. First, we tested for the main effect of the treatment on student science learning using the model above (β_3). Next, we tested for possible moderator effects associated with various student characteristics (prior mathematics and English Language Arts achievement, gender, race/ethnicity) by adding an interaction term (treatment indicator by student characteristic) at the school level. Separate models were run for each moderator variable by adding the corresponding interaction term for each potential moderator. We used the following groups in the moderator analysis: male versus female, Hispanic versus non-Hispanic, and white versus non-white. In addition, we created two subgroups based on the student 6th grade performance on ELA and mathematics achievement tests: at or above the median and below the median.

Exploratory Analysis: Impact on Teacher Instructional Practice. We also conducted an exploratory analysis to investigate whether differences in instructional practice in the treatment group may help explain any impacts found of the AS curriculum program on student learning (RQ 2b). We consider this analysis *exploratory* because the measures of instructional practice are based on teacher self-reported data and the sample size is relatively small (32 teachers from 3 districts and 18 schools). Because of the small sample size, a single-level regression model was used to estimate the treatment and control differences in teacher instructional practice. A computation of the standard error of the estimated impacts that allows for intragroup correlation was used to account for the clustering effect of the data (teachers nesting within schools). Single-item science instruction measures based on data collected from the instructional logs were used as the dependent variables. To improve the precision of the estimates a set of covariates were added at the teacher-level including the number of years teaching, number of years teaching middle school, number of years teaching science, science teaching certification (Yes or No), percent of students who were ELL across all class sections taught by the teacher, percent of students with an IEP, and the percent of students that were Asian, Hispanic, and White.

Results and Findings

Sample

As described above, the analytical sample consisted of 18 schools (10 treatment and 8 control) representing those schools and classrooms where teachers were able to complete their physical science lessons and the administration of the physical science assessment. A total of 33 teachers (19 treatment and 14 control) taught 7th grade science in these schools to a total of 3,738 students (1,940 treatment and 1,798 control). Student enrollment data is based on classroom rosters collected after the schools were randomized and within the first several weeks after the start of the school year.

Cluster Attrition and Individual Non-response

With 18 of the original 29 schools remaining in the analytic school sample, the overall cluster attrition rate was 37.9% with a differential attrition of 9.5% (33.3% for treatment schools compared to 42.9% for control schools). There was also a relatively high rate of student non-response primarily due to the timing of COVID-related school closures relative to when individual teachers completed their physical science instruction. To estimate the percent of student non-response, we first analyzed the number of students who were enrolled in 7th grade in a participating teachers' classrooms in fall 2019 and remained in the study (stayers). Of the 3,738 students who appeared on the classroom rosters of the 18 schools in the analytic sample in the fall of 2019, we received parental consent and student assent from 2,457 students (1,235 treatment and 1,222 control). Of these 2,457 students, 1,953 remained enrolled in their schools during the study and completed the study's physical science assessment (913 treatment and 1,040 control). Of the 1,785 student non-responders, 1,027 were in the treatment condition and 758 were in the control condition. Thus, the resulting overall individual non-response rate was 47.8% (52.9% for the treatment group and 42.2% for the control group) with a differential non-response rate between treatment and control groups of 10.8%.

Due to the high level of cluster attrition and individual non-response there is a potential threat of bias due to compositional changes in the reference sample following randomization (What Works Clearinghouse, 2022). To establish the extent of the possible bias due to compositional changes in the student sample, we tested next for the equivalence between the experimental conditions on a set of baseline measures of prior academic achievement and student demographics.

Baseline Equivalence of Analytic Sample

Table 3 shows the prior achievement scores and demographics for students in the analytic sample at the start of the school year by experimental condition. The analytic sample was 41% White, 23% Hispanic, 21% Asian, 4% Black, and 11% Other ethnicity. Ten percent of the student sample had an individualized education program indicating that these students potentially received some type of specialized instruction and related services to supplement their science instruction. Only 5% of the sample had an English language learner (ELL) designation. We found small differences between conditions for prior (6th grade) standardized ELA test scores and for mathematics, corresponding to effect sizes (Hedges g) of −.11 and −.06 standard deviation units respectively, favoring the control group. In addition, we also found statistically significant differences between conditions in ethnic composition by condition with students in the treatment schools more likely to be White (47% versus 37%) and control students more likely to be Asian (26% vs. 16%) and Hispanic (26% vs. 16%). Since the differences in baseline prior achievement scores are less than .25 standard deviation units but greater than .05, we statistically control for these differences in our estimation of the impacts of being assigned to the AS curriculum by including measures of prior ELA and mathematics achievement and ethnicity in the analytical models (What Works Clearinghouse, 2022).

Treatment Teachers Implementation of the AS Curriculum (RQ 2a)

Teachers in the treatment group reported that they implemented a strong majority of the lessons and activities that were available in the AS curriculum for the teaching of physical science. Approximately 98% of the time a treatment teacher reported in the instructional logs that they taught a topic in physical science, they also reported they implemented one or more lessons or activities in the AS curriculum to teach that topic. In total there were 38 lessons and activities that teachers could have implemented, distributed evenly across the two units. Treatment teachers reported that they implemented 87% of the lessons and activities provided in the curriculum for the teaching of physical science, including 81% of the lessons and activities for the Phase Change unit and 93% for the Chemical Reactions unit. Noteworthy is that when teaching the two physical science units, treatment teachers also reported that, on average, they modified or supplemented 23% of the AS lessons and activities in some manner. Teachers reported they made modest modifications for a variety of reasons including to better meet the needs of their students and to make sure they completed the lesson or activity in the instructional time available.

Main Impact Analysis (RQ 1a and 1b)

Results from the analysis of the estimated impact of assignment to the AS curriculum condition on student

TABLE 3

Mean and standard deviation of demographic characteristics and prior test scores for students in the final analytical sample along with results of tests for baseline equivalence

	All		Treatment		Control		
Variables	Mean (N)	Standard Deviation	Mean (N)	Standard Deviation	Mean	Standard Deviation	Difference (T–C)
Female	.47 (1945)	.50	.48 (909)	.50	.51 (1036)	.50	−.04
English language learner status	.05 (1586)	.23	.06 (663)	.24	.05 (923)	.22	+.01
IEP status	.10 (1586)	.29	.10 (663)	.29	.10 (923)	.30	+.00
Asian[a]	.21 (1946)	.41	.15 (909)	.36	.26 (1037)	.44	−.11***
Black[a]	.04 (1946)	.19	.04 (909)	.20	.04 (1037)	.19	.00
Hispanic[a]	.23 (1946)	.42	.22 (909)	.42	.24 (1037)	.43	+.02
White[a]	.41 (1946)	.49	.49 (909)	.50	.34 (1037)	.48	+.14***
Others[a]	.11 (1946)	.31	.10 (909)	.30	.11 (1037)	.32	−.02
Grade 6 Math (standardized)[b]	.56 (1775)	.96	.52 (842)	.97	.57 (933)	.95	−.06
Grade 6 ELA (standardized)[b]	.56 (1767)	.96	.50 (836)	.95	.61 (931)	.97	−.10

***$p < .001$.
[a]The standard deviation for each ethnic group was computed based on a binary indicator for each ethnic group. For example, for Asians, Asian versus non-Asians.
[b]The reported means for Grade 6 mathematics and ELA are regression-adjusted means.

TABLE 4

Results of two-level hierarchical regression model to estimate main effect of treatment (AS curriculum)

	Adjusted Mean					
Measure (0–25 scale)	Treatment N = 913	Control N = 1040	Adjusted Mean Difference	Standard Error	P-value	Effect Size (Hedges g)
Assessment of NGSS Physical Science Performance Expectations (Matter and Its Interactions; Total Score)	13.773	11.756	+2.071	.4084	<.001	.40

learning show that students in the treatment schools scored 8.1% higher (2.02 points higher on a 0–25 point scale) on the physical science assessment than did students in the comparison schools (see Table 4). The results were similar across gender and racial and ethnic groups and for students with different prior math and literacy achievement (see Table 5). The estimated impact was statistically significant (p < .001) and corresponds to an effect size of .40 (Hedges' g). This effect size is equivalent to the average student in the treatment schools improving 16 percentile points (moving from 50th to 66th percentile) relative to the average student in the control schools (What Works Clearinghouse, 2022).

Exploratory Analysis: Impact on Teacher Instructional Practice (RQ 2b)

The results for the exploratory analysis investigating possible differences in the instructional practices by experimental condition are shown in Tables 6 and 7. Due to the small number of teachers in the analytic sample, in our summary

TABLE 5

Results of two-level hierarchical regression models to identify potential moderators of treatment effect (subgroup analyses)

	Adjusted Mean (0-25 Scale)				Adjusted Mean Difference	Contrast (Interaction)	Standard Error	P-value	Effect Size (Hedges g)
	Treatment	(N)	Control	(N)					
Female	14.032	434	11.749	532	2.283	.5195	.3373	.123	+.10
Male	13.531	475	11.768	504	1.764				
Hispanic	13.115	203	11.396	253	1.719	−.4018	.4191	.338	−.08
Non-Hispanic	14.054	623	12.104	765	1.950				
White	13.981	444	11.957	357	2.024	.0118	.3632	.974	.00
Non-White	13.628	469	11.615	683	2.012				
Grade 6 Math (high)	14.859	402	12.984	485	1.875	−.5694	.4024	.157	−.11
Grade 6 Math (low)	13.073	440	10.629	448	2.444				
Grade 6 ELA (high)	14.723	393	12.616	485	2.107	.0585	.4425	.890	+.01
Grade 6 ELA (low)	13.233	443	11.184	446	2.049				

TABLE 6

Results of linear regression models to estimate treatment effect on the average percent of instructional time devoted to NGSS science and engineering practices

Measure (6-point scale, 1 to 6, Not at all to More than 75%)	Adjusted Mean		Adjusted Mean Difference	P-value	Effect Size (Hedges g)
	Treatment N=18	Control N=14			
Students ask science questions	2.982	3.563	−0.581	.060	−0.46
Students develop and use models	3.461	3.469	−0.008	.977	−0.01
Students plan and carry out investigations	1.702	3.335	−1.634	<.01	−1.54
Students analyze and interpret data	2.927	3.481	−0.553	.015	−0.59
Students use mathematics and computational thinking	1.387	2.324	−0.937	.005	−1.28
Students construct explanations	3.493	3.390	0.103	.666	+0.10
Students engage in argument from evidence	3.139	2.839	0.300	.336	+0.33
Students obtain, evaluate, and/or communicate information	3.638	3.312	0.326	.252	+0.31

TABLE 7

Results of linear regression models to estimate treatment effect on the percent of weekly logs reporting teacher provided instructional opportunity

Measure (% of weekly logs)	Adjusted Mean		Adjusted Mean Difference	P-value	Effect Size (Hedges g)
	Treatment N=18	Control N=14			
A hands-on experience that supported science learning	.588	.720	−.132	.011	−0.61
A demonstration for students to watch	.294	.419	−.125	.024	−0.57
Opportunities for students to do individual seat work	.908	.810	.098	.437	+0.51
Opportunities for students to work collaboratively	.927	.924	.003	.929	+0.03
Opportunities for students to read science text	.776	.469	.307	.091	+1.38
Opportunities for students to connect science learning to everyday experiences	.724	.616	.108	.360	+0.40
Opportunities for students to write in science notebooks (e.g., writing in paper or digital notebooks or within digital platforms where the writing is archived)	.890	.730	.161	.030	+0.87
Opportunities for students to communicate their scientific thinking to one another	.890	.774	.116	.019	+0.73
A quiz for students to complete	.351	.318	.033	.784	+0.15

of the results below we highlight differences in the self-reported science instruction measure between treatment and control schools that are generally considered moderate to large effects for education research (greater than or equal to .20 standard deviations), whether they are statistically significant or not (e.g., Hattie, 2008; Hill et al., 2008). Differences of this magnitude represent promising evidence of the way the use of the AS curriculum and professional learning may have impacted science instruction in treatment classrooms during the study.

Coverage of NGSS Science and Engineering Practices. The results for measures of the coverage of the eight NGSS science practices are presented in Table 6. The results provide us with some insight into how the intervention may have changed the instructional environment in treatment school by comparing the extent to which teachers in both conditions engaged their students with opportunities to engage in the eight NGSS science and engineering practices during instruction of physical science units. Table 6 shows that science teachers in treatment schools self-reported devoting more instruction time to two of the eight scientific practices (effect size ≥ 0.20 stand deviations) compared to teachers in the control schools. Treatment teachers were more likely than control teachers to report they spend instructional time engaging students in *engaging in argument from evidence* (effect size $= +0.33$, p $= .336$) and *obtaining, evaluating and/or communicating information* (effect size $= +0.31$, p $= .252$). In contrast, control teachers were more likely than treatment teachers to report they spend instructional time engaging students in *planning and carrying out investigations* (effect size $= -1.54$, p $< .001$; *using mathematics and computational thinking* (effect size $= -1.28$, p $= .005$); *analyzing and interpreting data* (effect size $= -0.59$, p $= .015$); and *asking science questions* (effect size $= -0.46$, p $= .060$).

Prevalence of Instructional Opportunities. The results for measures of the prevalence of the nine instructional opportunities are presented in Table 7. These results provide a glimpse into how the intervention may have changed the instructional environment in treatment classrooms across a set of activities that represent a range of independent, collaborative, passive and active learning activities. Table 7 shows that treatment teachers were more likely than control teachers to self-report they provided students with opportunities to *read scientific texts* (effect size $= +1.38$, p $= .091$); *write in science notebooks or the digital equivalent* (effect size $= +0.87$, p $= .030$); *communicate their scientific thinking to one another* (effect size $= +0.73$, p $= .019$); *do individual seat work* (effect size $= +0.51$, p $= .437$); and *connect science learning to everyday experiences* (effect size $= +0.40$, p $= .360$). Compared to treatment teachers, control teachers reported providing more opportunities for students to engage in *hands-on experience*s (effect size $= -.61$, p $= .011$) and *demonstrations for students to watch* (effect size $= -.57$, p $= .024$).

Discussion

In this study, we found that the use of the AS curriculum materials improved student learning in the context of physical science instruction that aimed to help students' build proficiency toward NGSS performance expectations. The curriculum was designed to promote student engagement with SEPs, DCIs, and CCCs through online simulations; reading, writing and drawing tasks to explain phenomena; and through select hands-on activities. At posttest, the treatment students scored significantly higher than the control group on the physical science topics. The main effect we observed is relatively large and greater than what has been reported in prior experimental studies of science curriculum materials (Harris et al., 2015; Krajcik et al., 2023; Taylor et al., 2015; Wilson et al., 2010).

In our exploratory analysis of curriculum implementation, we found that teachers in the treatment condition adhered to the main instructional activities of the curriculum, though many reported that they made some adaptations. Weekly instructional logs completed by both treatment and control teachers made it clear that there were differences in the types of instructional opportunities provided and in the emphasis on SEPs. Regarding science instruction and the dimensions measured, we found that the substantive support for science and literacy integration within the curriculum was taken up by treatment teachers and students and accomplished via reading, writing, and discourse activities that frequently leveraged the SEPs of engaging in argument from evidence and obtaining, evaluating and communicating information. Teachers in the control condition reported that they engaged students in hands-on activities, investigations, and demonstrations more often. In end-of-year surveys and interviews, teachers using the Amplify Science curriculum reported that this was a perceived shortcoming of the materials—they would have preferred more hands-on investigations in the unit lessons. Teachers using AS also reported more seatwork where students were more likely to be interacting with the digital curriculum and the online simulations.

The finding that the teachers in the treatment group spent less time with their students on planning and conducting scientific investigations is indicative of what is emphasized less in the NGSS-designed AS materials. At this time, there is very little research evidence available regarding what may be the best weighting of SEPs that should be included within instructional sequences in curricula. What is clear is that the emphasis on scientific practices vary in the types of curriculum materials currently available including NGSS-designed curriculum materials that are going to scale. Planning and carrying out investigations is central to science and research

on classroom science investigations suggest that conducting them has many learning benefits for students (National Academies of Sciences Engineering and Medicine [NASEM], 2019). For example, as students investigate questions, pursue solutions, and work together on investigating phenomena, they are afforded many opportunities to actively think about, integrate, and apply ideas and thereby deepen their science proficiency (Windschitl, 2017). Yet, orchestrating them to fully realize those benefits can be challenging at times in a classroom environment (NASEM, 2019). Challenges can arise for teachers in monitoring, supporting, and sustaining students' progress in their investigations. When pacing is thrown off and investigations extend too long, there can be little or no time remaining at the end of lessons for teachers to debrief with students, facilitate discourse on the phenomena under study, and fully address the important science ideas with their students (Harris & Rooks, 2010). This is an all-too common occurrence in science classrooms. Our study did not include observations of the investigations conducted in classrooms and thus we are not able to provide further insight. More work remains to be done on the benefits and trade-offs of time devoted to the various SEPs and specifying the situations in which they are most likely to be effective in advancing student learning.

Importantly, this study also draws attention to the critical role of efficacy research in building an evidence base for NGSS-designed curricula. The Framework and the NGSS expand perspectives on science proficiency from primarily what students know to also include how they can use and apply what they know to make sense of phenomena and design solutions to problems. This vision for learning has required changes in how science curriculum and instruction are conceptualized, supported, and implemented. Put broadly, the results show that curriculum materials expressly designed for NGSS teaching and learning, along with accompanying professional learning, can support educators in creating classroom conditions that will prepare students for next generation science learning. This is important because it highlights that highly specified and developed curriculum materials matter for shifting classroom practice toward the vision of the Framework and the performance expectations of the NGSS. These results also direct our attention to the need for further research on the design principles and features within materials that may account for increased learning.

Lastly, in this study we defined the Amplify Science curricular program as an approach to teaching science that includes instructional resources (e.g., lesson plans and unit guides), teacher professional learning, student materials, and technology. In a synthesis of research on 6-12 grade science programs, Cheung and colleagues (Cheung et al., 2017) found that programs with applications of technology to support science instruction and with teacher professional learning to equip teachers for implementation have been more

successful in improving learning as evidenced in experimental evaluations. The findings in this study lend support to what Cheung and colleagues and others have found regarding the types of programs that are likely to elevate teaching and make a difference for students' science learning.

Limitations of Study

There are several limitations to this study. A first limitation is the attrition of schools due to school closures at the onset of the COVID-19 pandemic. At the time of school closures, participating schools were in the midst of completing instruction and administering assessments in physical science. Only schools that fully completed their physical science instruction and administered the post-assessment were included in the learning outcome analysis. Accordingly, the overall and differential school-level attrition and student non-response rate were relatively high which raises some concern that this might introduce bias into the study's results. However, our test of baseline equivalence of the analytic sample on prior achievement in reading and mathematics found no evidence of significant bias; in fact the differences we found favored the control group, which would indicate our estimates of the impacts of the AS curriculum on learning might be understated.

A second limitation is that this study focused on just one domain and at one grade level that is part of the more comprehensive AS curriculum spanning across the 6–8 middle grades. The full curricular intervention covers all the science domains and is designed to support instruction toward meeting the breadth of NGSS performance expectations for this grade band. Had the study included more domains and grade levels, a more definitive conclusion could have been drawn regarding the overall impact on learning for the curriculum. Still, because the full range of AS units that span the domains are infused with the same pedagogical approach and were developed with the same design principles (Loper et al., 2022), the study's findings on student learning in the domain of physical science stand to provide evidence of promise for the overall NGSS-designed curriculum. Importantly, the findings encourage further research on AS curriculum implementation and its impact on student learning at other grade levels and within other science domains.

A third potential limitation of the study is that the teachers in the two conditions did not receive an equal amount of professional learning during the course of the study that was customized to their respective curriculum materials. Teachers in the treatment condition were new to Amplify Science and received the standard professional learning that is part of the launch of the program. In contrast, teachers in the control condition were experienced with their district adopted materials and had access to whatever science professional learning was available to them in their districts. In the end, teachers in the control condition did not receive an

equivalent amount of professional development customized to their business-as-usual curricula during the study. Yet, it is important to note that, prior to the study, teachers in both conditions had taught science in their schools for an average of more than 4 years (4.44 years for treatment teachers compared to 4.43 for control teachers). Thus, during the study, it is highly likely that the average teacher in the control condition was as or more familiar with the implementation of their curriculum materials ("business-as-usual") than the average teacher in the treatment condition who was using the AS materials for the first time. Thus, we do not consider the amount of customized professional learning provided during the study to be a potential confounding factor that might lead to a spurious association between teachers' uptake of the AS curriculum program and improved student learning.

Finally, a fourth limitation of the study was related to the study's reliance on teacher self-reports for measures of the instructional environment. Due to school closures, the study team was not able to implement the planned classroom observations to independently validate the nature of the classroom instruction in treatment and control classrooms and implementation of the AS curricula in treatment classrooms. Instead, we had to rely on teachers' own self-reports of the instructional environment using weekly instructional logs during the teaching of physical science in tandem with an end-of-the-year survey. Consequently, we must be cautious in our interpretation of the exploratory analysis of the data from the instructional measures as the use of self-reported measures may introduce bias that might influence the analytical results (Althubaiti, 2016).

Conclusion

To date, few NGSS-designed science curriculum programs have been rigorously examined. The results from this randomized controlled trial contribute to this emerging research base. This study joins a small yet increasing number of experimental studies extending from the elementary grades (e.g., Harris et al., 2023; Krajcik et al., 2023), to middle school (e.g. Harris et al., 2015) and to high school (e.g., Schneider et al., 2022) that are examining classroom implementation and impact on learning of curriculum materials that aim to support today's vision for science education. We are in an era where new programs are becoming widely available across K–12. Among recent examples are the OpenSciEd materials for K–12 science (Edelson et al., 2021) and the elementary-based Collaborate Science (formerly known as Multiple Literacies in Project-Based Learning) program (Krajcik & Schneider, 2021) that are now being released widely. As new programs become more broadly taken up across different geographic regions and with varying student populations, additional studies at larger scale and with concerted attention to student diversity and equity will

be needed. This current and future research work is greatly needed and will be critical for ensuring that the vision of the Framework and the NGSS is realized for all students.

ORCID iDs

Robert F. Murphy 🔟 https://orcid.org/0000-0001-8747-9816
Kevin Huang 🔟 https://orcid.org/0009-0005-3855-7036

Notes

1. This study was conducted by an independent team of researchers from WestEd, SRI Education, and LFC Research. Developers from the University of California, Berkeley's Lawrence Hall of Science provided the professional learning support that was part of the curricular intervention. The Lawrence Hall of Science also served as a resource and thought partner to the study team on matters related to the curriculum, its design features, and its implementation. Amplify Education, Inc., which is the publisher and distributor of Amplify Science, was not involved in the administration of the study. All data collection and analysis activities were carried out independently by the study team and did not involve representatives of the publisher or the Lawrence Hall of Science.

2. We first calculated the Mahalanobis distance between all schools within a block based on the four matching variables, then conducted 2000 random pairing configurations between schools within a block and selected the pairing configuration with the lowest average Mahalanobis distance across pairs.

3. One group of school matchings involved a three-school triad. In this case, two schools were randomly assigned to the treatment condition and one to control.

4. Technology-enhanced versions of the tasks that were adapted for this study, among other NGSS-focused tasks, can be found on the publicly accessible Next Generation Science Assessment Task Portal, https://ngss-assessment.portal.concord.org/middle-school.

References

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, *9*, 211–217. https://doi.org/10.2147/JMDH.S104807

Anderson, C. W., de los Santos, E. X., Bodbyl, S., Covitt, B. A., Edwards, K. D., Hancock, J. B., Lin, Q., Thomas, C. M., Penuel, W. R., & Welch, M. M. (2018). Designing educational systems to support enactment of the next generation science standards. *Journal of Research in Science Teaching*, *55*(7), 1026–1052. https://doi.org/10.1002/tea.21484

Cheung, A., Slavin, R. E., Kim, E., & Lake, C. (2017). Effective secondary science programs: A best-evidence synthesis.

Journal of Research in Science Teaching, *54*(1), 58–81. https://doi.org/10.1002/tea.21338

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

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2016). Building an assessment argument to design and use next generation science assessments in efficacy studies of curriculum interventions. *American Journal of Evaluation*, *37*(2), 174–192. https://doi.org/10.1177/1098214015581707

Duncan, R. G., Krajcik, J., & Rivet, A. E. (2017). *Disciplinary core ideas: Reshaping teaching and learning*. National Science Teachers Association Press.

Edelson, D. C., Reiser, B. J., McNeill, K. L., Mohan, A., Novak, M., Mohan, L., Affolter, A., McGill, T. A. W., Buck Bracey, Z. E., Deutch Noll, J., Kowalski, S. M., Novak, D., Lo, A. S., Landel, C., Krumm, A., Penuel, W. R., Van Horne, K., González-Howard, M., & Suárez, E. (2021). Developing research-based instructional materials to support large-scale transformation of science teaching and learning: The approach of the OpenSciEd middle school program. *Journal of Science Teacher Education*, *32*(7), 780–804. https://doi.org/10.1080/1046560X.2021.1877457

EdReports. (2020). Review of science instructional materials: Amplify Science 6–8. https://edreports.org/reports/overview/amplify-science-2018

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

Gu, X., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, *2*(4), 405–420. https://doi.org/10.1080/10618600.1993.10474623

Harris, C. J., Krajcik, J., Pellegrino, J., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice*, *38*(2), 53–67. https://doi.org/10.1111/emip.12253

Harris, C. J., Murphy, R., Feng, M., & Rutstein, D. (2023). *Supporting science learning and literacy development together: Initial results from a curriculum study in 1st grade classrooms*. WestEd.

Harris, C. J., Penuel, W. R., D'Angelo, C., DeBarger, A. H., Gallagher, L. P., Kennedy, C., Cheng, B. H., & Krajcik, J. S. (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. https://doi.org/10.1002/tea.21263

Harris, C. J., Phillips, R. S., & Penuel, W. R. (2012). Examining teachers' instructional moves aimed at developing students' ideas and questions in learner-centered science classrooms. *Journal of Science Teacher Education*, *23*(7), 769–788. https://doi.org/10.1007/s10972-011-9237-0

Harris, C. J., & Rooks, D. L. (2010). Managing inquiry-based science: Challenges in enacting complex science instruction in K–8 classrooms. *Journal of Science Teacher Education*, *21*(2), 227–240. https://doi.org/10.1007/s10972-009-9172-5

Hattie, J. (2008). *Visible learning*. Routledge.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Huffman, D., Goldberg, F., & Michlin, M. (2003). Using computers to create constructivist learning environments: Impact on pedagogy and achievement. *Journal of Computers in Mathematics and Science Teaching*, *22*, 151–168.

Institute of Education Sciences & National Science Foundation. (2013). *Common guidelines for education research and development*. https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf

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

Krajcik, J., & Schneider, B. (2021). *Science education through multiple literacies: Project-based learning in elementary school*. Harvard Education Publishing Group.

Krajcik, J., Schneider, B., Miller, E. A., Chen, I. C., Bradford, L., Baker, Q., Bartz, K., Miller, C., Li, T., Codere, S., & Peek-Brown, D. (2023). Assessing the effect of project-based learning on science learning in elementary schools. *American Educational Research Journal*, *60*(1), 70–102. https://doi.org/10.3102/00028312221129247

Lee, O., Deaktor, R., Enders, C., & Lambert, J. (2008). Impact of a multiyear professional development intervention on science achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, *45*(6), 726–747. https://doi.org/10.1002/tea.20231

Loper, S., Barber, J., Greenwald, E., Krakowski, A., Morales, C., Allan, S., Abbott, R., Kitagawa, L., Montgomery, R., Ouyang, A., Pastor, M. J., Stenger, L., & the Learning Design Group. (2022, September 26–29). *Designing curriculum to support deep, equitable science learning at scale: Amplify Science design goals and design features* [Poster presentation]. The Annual Meeting of the International Society for Design and Development in Education, Nottingham.

Lowell, B. R., Cherbow, K., & McNeill, K. L. (2021). Redesign or relabel? How a commercial curriculum and its implementation oversimplify key features of the NGSS. *Science Education*, *105*(1), 5–32. https://doi.org/10.1002/sce.21604

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (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. https://doi.org/10.1002/tea.20039

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

National Academies of Sciences Engineering and Medicine. (2019). *Science and engineering for grades 6–12: Investigation and design at the center*. The National Academies Press.

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

National Research Council. (2014). *Developing assessments for the Next Generation Science Standards*. National Academies Press.

National Research Council. (2015). *Guide to implementing the Next Generation Science Standards*. National Academies Press.

National Science Teaching Association. (2022). *NGSS@NSTA: STEM starts here*. https://ngss.nsta.org/

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Nordine, J., & Lee, O. (Eds.). (2021). *Crosscutting concepts: Strengthening science and engineering teaching and learning*. National Science Teachers Association.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*, 459–463. https://doi.org/10.1126/science.1182595

Pellegrino, J. W., Wilson, M., Koenig, J., & Beatty, A. (Eds.). (2014). *Developing assessments for the Next Generation Science Standards*. 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. https://doi.org/10.3102/0002831211410864

Penuel, W. R., & Reiser, B. J. (2018). *Designing NGSS-aligned curriculum materials*. Committee to Revise America's Lab Report. National Academies of Science, Engineering, and Medicine.

Roblin, P. N., Schunn, C., & McKenney, S. (2018). What are critical features of science curriculum materials that impact student and teacher outcomes? *Science Education*, *102*, 260– 282.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*, 573–585. https://doi.org/10.1080/01621459.2000.10474233

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., Chen, I.-C., Baker, Q., Touitou, I., Peek-Brown, D., Dezendorf, R., Maestrales, S., & Bartz, K. (2022). Improving science achievement—Is it possible? Evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educational Researcher*, *51*(2), 109–121. https://doi.org/10.3102/0013189X211067742

Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). *Helping students make sense of the world using next generation science and engineering practices*. NSTA Press, National Science Teachers Association.

Short, J., & Hirsh, S. (2020). *The elements: Transforming teaching through curriculum-based professional learning*. Carnegie Corporation of New York.

Smith, P. S. (2020). *Obstacles to and progress toward the vision of the NGSS*. Horizon Research.

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*, 984–1017. https://doi.org/10.3102/0002831215585962

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, Version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE).

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and common-place science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, *47*(3), 276–301. https://doi.org/10.1002/tea.20329

Windschitl, M. (2017). Planning and carrying out investigations. In C. V. Schwarz, C. Passmore, & B. J. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 135–157). NSTA Press.

Authors

ROBERT F. MURPHY is an independent consultant and principal education researcher and program evaluator. He designs and leads experimental, quasi-experimental, and mixed-method studies of innovative programs and technologies designed to support teaching and learning.

CHRISTOPHER J. HARRIS serves as senior director of Science and Engineering Education Research at WestEd. His research focuses on the design and study of curricular, instructional, and assessment innovations that support science teaching and learning in PK–12 classrooms.

MINGYU FENG is a research director with WestEd's Learning and Technology team. She leads large-scale studies focused on leveraging education technologies to transform science and mathematics instruction to improve student learning.

ASHLEY IVELAND is a senior research associate with the Science and Engineering team at WestEd. She leads multiple large-scale projects developing and studying teacher professional learning focused on equitable and context-driven science and engineering teaching and learning at the elementary and middle school levels.

MELISSA REGO is a research associate with the Science and Engineering team at WestEd. She specializes in qualitative data collection, analysis, and project coordination for a variety of large- and small-scale research studies in primary and secondary STEM education.

DAISY RUTSTEIN is a senior associate at edCount, LLC. She specializes in applying an evidence-centered design process to the design, development and implementation of learning assessments.

KEVIN HUANG is a senior research associate at WestEd. He applies quantitative and statistical methods to support the design, implementation, and data analysis of rigorous experimental trials as well as measurement and assessment projects.