

Connecting the Dots from Professional Development to Student Learning

Charlene L. Ellingson,[†] Katherine Edwards,[‡] Gillian H. Roehrig,[§]
M. Clark Hoelscher,^{||} Rachelle A. Haroldson,[¶] and Janet M. Dubinsky^{**}

[†]Department of Elementary and Literacy Education, Mankato State University, MN 56001;
[‡]Minnesota Department of Education, Roseville, MN 55113; [§]STEM Education Center, University of
Minnesota, St. Paul, MN 55108; [¶]Department of Teacher Education, University of Wisconsin–River
Falls, WI 54022; ^{**}Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455

ABSTRACT

Following professional development (PD), implementation of contemporary topics into high school biology requires teachers to make critical decisions regarding integration of novel content into existing course scope and sequence. Often exciting topics, such as neuroscience, do not perfectly align with standards. Despite commitment to enacting what was learned in the PD, teachers must adapt novel content to their perceptions of good teaching, local context, prior knowledge of their students, and state and district expectations. How teachers decide to integrate curricula encountered from PD programs may affect student outcomes. This mixed-methods study examined the relationship between curricular application strategies following an inquiry-based neuroscience PD and student learning. Post-PD curricular implementation was measured qualitatively through analysis of teacher action plans and classroom observations and quantitatively using hierarchical linear modeling to determine the impact of implementation on student performance. Participation in neuroscience PD predicted improved student learning compared with control teachers. Of the two distinct curricular implementation strategies, enacting a full unit produced significantly greater student learning than integrating neuroscience activities into existing biology units. Insights from this analysis should inform teacher implementation of new curricula after PD on other contemporary biology topics.

INTRODUCTION

Teacher professional development (PD) is essential to support K–12 science education reform given the pedagogical changes called for in previous and current reform initiatives (National Research Council [NRC], 2012; Windschitl and Stroupe, 2017; Fischer *et al.*, 2018). PD research has typically focused on how participation augments teacher knowledge, beliefs, and changes in classroom practice (Marx *et al.*, 2004; Lee *et al.*, 2005; Lotter *et al.*, 2007; Blanchard *et al.*, 2009). The accepted theory of action for PD is that participation in quality PD causes improved teacher-level outcomes (e.g., content knowledge, beliefs, reform-based practices), which in turn leads to improved student outcomes (e.g., Roth *et al.*, 2019). However, relatively few studies report on the impact of PD using careful experimental or comparison group designs (Darling-Hammond *et al.*, 2017).

PD focused on the implementation of reform-based curricular materials further increases the impact of PD on teacher practices and student learning (Powell and Anderson, 2002; Schneider and Krajcik, 2002; Taylor *et al.*, 2015; Taylor *et al.*, 2018; Lynch *et al.*, 2019; Kowalski *et al.*, 2020). For example, Lynch *et al.* (2019) in a recent meta-analysis reported significantly larger effect sizes for student outcomes when both PD and curricular materials were provided to teachers rather than just PD or curricular materials. Curricular materials are important, because they serve as resources teachers draw upon to craft instruction (Brown and Edelson, 2003; Brown, 2009), supplying

Stephanie Gardner, *Monitoring Editor*

Submitted Feb 11, 2021; Revised Aug 02, 2021;

Accepted Aug 13, 2021

CBE Life Sci Educ December 1, 2021 20:ar57

DOI:10.1187/cbe.21-02-0035

^{||}Present address: Saint Paul Public Schools, 600 Weir Drive, Woodbury, MN 55125.

^{**}Address correspondence to: Janet M. Dubinsky (dubin001@umn.edu).

© 2021 C. L. Ellingson *et al.* CBE—Life Sciences Education © 2021 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

guidance in planning, providing background information on content, and specifying the scope and sequencing of instruction. While education reform documents delineate broad principles and guidelines for teachers, “curriculum and professional development provide direct instructional guidance, attempting to shape schools’ and teachers’ day-to-day interactions with students” (Lynch *et al.*, 2019, p. 261). However, little is known about how the nature of teachers’ implementation of curricular materials introduced during PD impact student outcomes. The local context must be examined through in-depth, classroom observations for evidence of the specific practices that produce student gains (Fischer *et al.*, 2018). Thus, a pressing concern for researchers is to establish direct links between participation in professional learning opportunities and student outcomes, considering how teachers implement curricula and practices they learned in the PD (Blank and de las Alas, 2009; Capps *et al.*, 2012). To address this gap in the literature, this study was guided by the following research questions:

1. What is the effect of participation in the PD on student learning?
2. How does the nature of the implementation of curricular materials from the PD impact student learning?

THEORETICAL FRAMEWORK

Curricular materials play a critical role in K–12 science classrooms, because they provide teachers with guidance for what content to teach and how to teach content (NRC, 2002; Roblin *et al.*, 2017), as well as strategies for differentiating instruction to improve student learning outcomes (Forbes and Davis, 2010; Taylor *et al.*, 2015). However, curricular materials themselves are inert. Teachers ultimately determine how curricular materials are used (Powell and Anderson, 2002; Roehrig *et al.*, 2007; Forbes and Davis, 2010) due to their subject matter and pedagogical content knowledge (Brown, 2009), their goals and beliefs about teaching and learning, and how the nature of the implementation impacts student learning (Roehrig and Garrow, 2007).

Conceptions of curricular materials and their use have shifted from conceiving of the teacher’s role as that of enactor of prescribed curricula toward conceiving of the teacher–curriculum relationship as “participatory” (Remillard, 2005, 2018; Brown, 2009), that is, there is a dynamic relationship between teachers and curricular materials (Brown and Edelson, 2003; Brown, 2009). New paradigms of curriculum use (Taylor, 2012) shift from the notion that curriculum is fixed and the teacher’s role is to implement it with fidelity toward a paradigm of teaching as a design activity (Brown and Edelson, 2003; Brown, 2009). Acknowledging teachers’ role in adapting PD resources and curricula promotes teachers’ autonomy, while fitting PD and curricular goals into the local context (Yoon and Klopfer, 2006). Unquestionably, teachers modify and adapt curricular materials to align with their perceptions of good teaching, state and district expectations, and the context of their classrooms (Durlak and DuPre, 2008; O’Donnell, 2008; Hill and Erickson, 2019). It can reasonably be expected that these local modifications may impact student learning outcomes.

Thus, a major debate in implementation research across many fields (e.g., education, healthcare) centers around fidelity and adaptation (O’Donnell, 2008). Teachers make important

decisions related to adherence to strategies promoted in PD and subsequent use of program materials (Hill and Erickson, 2019). While adaptation is inevitable and desirable to support student learning in context, the nature of the adaptations could “carry the risk of threatening the theoretical basis of the intervention, resulting in a negative effect on expected outcomes” (Perez *et al.*, 2016, p. 1). Treating curricula designed to promote inquiry-based learning as inflexible has been reported as problematic (Papaevripidou *et al.*, 2017). Implementation without adaptation provided for greater in-service teacher learning when the PD materials were novel (Knight-Bardsley and McNeill, 2016). Case study analysis of teachers’ lesson implementation decisions describe both favorable and unfavorable enactments. Favorable adaptations took into account students’ prior knowledge and abilities (Davis *et al.*, 2011) or existing difficulties (Skultety *et al.*, 2017) or responded to student questions (Sherin and Drake, 2009). Problematic adaptations arose from altered alignment between learning goals and curricular materials (Davis *et al.*, 2011; Knight-Bardsley and McNeill, 2016) and reduced student autonomy (Papaevripidou *et al.*, 2017).

The acknowledgment of the role of teachers in adapting curricula has led to the modifications of traditional fidelity frameworks to support the coexistence of fidelity and adaptation. Perez and colleagues’ (2016) modified fidelity framework argues for the systematic investigation of adaptations, including intervention components that are implemented as directed, not implemented, modified, or added by the user. Understanding the typology of adaptations is the first step to determining the impact of these adaptations on the intended outcomes. While the modified fidelity framework was developed to understand implementation of healthcare interventions, it has been used in the education field to understand curricular modifications made by educators (Bania *et al.*, 2017; Koster and Bower, 2018). Indeed, in a recent analysis of science, technology, engineering, and mathematics (STEM) PD studies, both moderate and high fidelity appear to yield similar likelihoods of positive student outcomes, suggesting that teachers do not have to rigidly enact a novel curriculum (Hill and Erickson, 2019). Teachers face a range of practical decisions regarding length of time to devote to a subject, which lessons best convey the content, and what practices best engage learners. Understanding the adaptive decisions teachers make may provide insights into what implementation strategies positively affect student outcomes (Hill and Erickson, 2019), rather than undercutting the potential of well-designed, coherent curricula (Davis and Varma, 2008; Penuel *et al.*, 2011).

LITERATURE REVIEW

National policy and standards documents over the past 25 years have called for changes in K–12 science teaching, emphasizing a shift from traditional teacher-centered instruction toward inquiry-based, student-centered approaches (NRC, 2002; Next Generation Science Standards [NGSS] Lead States, 2013). However, PD focused solely upon the adoption of curricula, science kits, or textbooks is largely ineffective at promoting these pedagogical changes as judged by student outcomes (Slavin *et al.*, 2014; Kleickmann *et al.*, 2016; Cheung *et al.*, 2017). The characteristics of PD programs needed to produce the changes in teacher practices advocated in policy documents

include: duration, content focused, collaborative, active learning with modeled pedagogy, opportunities for reflection, and follow-up support (Wilson, 2013; Darling-Hammond *et al.*, 2017). Unfortunately, there is a limited research base that links teachers' participation in PD guided by these characteristics to student learning outcomes designs (Darling-Hammond *et al.*, 2017). Within this review, 35 studies were identified that used rigorous research designs (experimental or quasi-experimental) to determine the impact of PD on student learning; only 12 of these studies were specific to science teaching.

While these 12 studies reported on 10 unique PD experiences, all with positive gains related to student learning, the studies represent a range of grade levels, PD focus, and research approaches. For example, while most PD programs focused on supporting teachers to implement inquiry-based approaches to teaching science, some focused on other pedagogical interventions, such as modifying instruction to be more culturally relevant (Johnson and Fargo, 2010, 2014), integrating reading strategies into high school biology classrooms (Greenleaf *et al.*, 2011), and integrating literacy strategies for English learners (Lara-Alecio *et al.*, 2012).

Most relevant to our study are PD programs that incorporated curricular materials to support teachers' implementation of inquiry-based teaching (e.g., Marek and Methven, 1991; Doppelt *et al.*, 2009; Penuel *et al.*, 2011; Roth *et al.*, 2011; Heller *et al.*, 2012; Kleickmann *et al.*, 2016; Taylor *et al.*, 2017). A critical component of effective PD is engaging teachers in the active-learning process situated in the curricula; simply providing teachers with curricula is significantly less impactful on student learning than when teachers have opportunities to engage in the curricular activities as learners (Doppelt *et al.*, 2009; Kleickmann *et al.*, 2016). Doppelt and colleagues (2009) explored the level of expert scaffolding within a PD necessary for subsequent curricular implementation to have a positive impact on student learning. The most successful PD implementation included expert scaffolding of both content and pedagogical content knowledge. The expert "challenged teachers' naive conceptions by engaging teachers in scientific investigations, giving (counter-) examples, prompting analogies, and stimulating discussions" (Doppelt *et al.*, 2009, p. 26). This scaffolding paralleled the scaffolding that would be necessary for elementary students engaging in the curricular activities, providing both pedagogical modeling and opportunities to reflect on the pedagogical needs of the curriculum.

Research related to PD supported by curricular materials acknowledges that curricular modifications are inevitable (e.g., Davis and Varma, 2008; Johnson and Fargo, 2010; Penuel *et al.*, 2011; Hill and Erickson, 2019). As argued in our theoretical framework, these modifications are inevitable and desirable to support student learning (Perez *et al.*, 2016). Many successful PD programs encourage and support curricular adaptations, providing professional learning for teachers to modify curricula to better meet the needs of a diverse student population (e.g., Johnson and Fargo, 2010, 2014; Penuel *et al.*, 2011). For example, Johnson and Fargo (2010) allocated time in the PD schedule for teachers to "reflect on and make modifications to the Premier Science curriculum to better meet the diverse needs of their students" (p. 13), work that was explicitly supported by professional learning related to culturally responsive teaching. Penuel and colleagues (2011) argue that PD models should

anticipate teachers' use of curricular materials and thus it is critical that teachers received explicit instruction in the models of teaching underpinning the curriculum. While actively engaging teachers in the inquiry-based, curricular activities to be enacted in the classroom is a widely agreed upon PD practice (Desimone, 2011; Wilson, 2013; Darling-Hammond *et al.*, 2017), providing explicit instruction in pedagogical approaches and models of inquiry-based learning remains important (Penuel *et al.*, 2011). In developing an understanding of the central model supporting the curricular design, teachers are supported in making useful, rather than problematic, curricular modifications that support student learning.

The impact of a PD is also influenced by external factors such as standardized testing and state, district, and school policies (Powell and Anderson, 2002; Roehrig *et al.*, 2007). Buczynski and Hansen (2010) explicitly explored barriers to implementing PD strategies in their study of upper elementary science PD. They argue that these external barriers, rather than teachers' beliefs and knowledge, make it difficult to implement practices promoted by the PD. These hurdles to PD implementation are most problematic in high-need urban schools (Johnson and Fargo, 2010). These external barriers also impact teachers' modification and use of curricular materials: "If teachers perceive the learning goals of curriculum materials to be incongruent with district or state standards, teachers may implement materials to a limited degree" (Penuel *et al.*, 2011, p. 999).

The increased emphasis on standardized testing and content coverage has led to a biology curriculum, at the middle and high school levels, that is "a mile wide and an inch deep" (Schmidt *et al.*, 1997, p. 122; Wood, 2002, p. 125). Biology educators send mixed messages to policy makers and teachers arguing that a high school course does not need to be all-inclusive, when, at the same time, research suggests that "when essential topics are omitted students miss opportunity to relate these topics to a broader perspective" (Vazquez, 2006, p. 31). Wood lists a series of current interest areas in biology not included in the curriculum and then goes on to state: "The principal problem, especially for AP courses, is not that they teach too little but that they attempt to teach too much" (Wood, 2002, p. 125). This problem is exacerbated when novel contemporary biology topics are added to engage and motivate students. As a result, the extensive list of high school biology standards generates a surface-level treatment of content with a focus on facts rather than conceptual understanding.

Indeed, the new frameworks and subsequent NGSS attempt to focus instruction on disciplinary core ideas, practices, and concepts considered central to each discipline (Bybee, 2013). This effort is supported by research that assesses student learning related to breadth and depth of content coverage. High school students who covered at least one major topic in depth for a month or longer earned higher grades in college biology than did students who reported no in-depth coverage (Schwartz *et al.*, 2009). In addition, students experiencing surface-level breadth in their high school biology courses, covering all major topics, were at a significant disadvantage in college-level biology courses (Schwartz *et al.*, 2008). However, criticisms of the NGSS documents include the critique that fundamental content is missing or vaguely stated, forcing the teacher to determine what content is necessary (Gross *et al.*, 2013). For example, a

review of genetics concepts within NGSS reported that Mendelian genetics, complex trait genetics, gene expression and regulation, and mutations as the basis for genetic disease were either lacking in coverage or completely absent (Lontok *et al.*, 2015). In the absence of strong curricular materials, life science teachers are forced to unpack the broad disciplinary core ideas to determine what content is to be taught. This process is complex, and recent research reports that teachers experience significant tension balancing content details and practices (Friedrichsen and Barnett, 2018). Even within the NGSS era, biology teachers' decisions about how to implement curricula promoted within a PD are complex in response to continued pressure for content coverage.

Additionally, many exciting contemporary biology topics that would produce student interest in STEM do not necessarily fit into the NGSS, including epidemiology (Dumais and Hasni, 2009), stem cell biology (Salli *et al.*, 2007), neuroscience (Gage, 2019), genetics (Yang *et al.*, 2017), gene editing/synthetic biology (Stark *et al.*, 2019), bioinformatics (Kovarik *et al.*, 2013), and integrating interdisciplinary topics like the effects of climate change on human health (Nagle, 2013). Trying to engage students by teaching contemporary biology poses problems of how to integrate such topics into the existing content standards. The focus on high-stakes testing may limit coverage of contemporary topics relevant to promoting student interest (Romine *et al.*, 2016). While some curricula addressing contemporary biological topics align to NGSS disciplinary core ideas, most only align to science and engineering practices or crosscutting concepts (Straus and Chudler, 2016), making them harder to integrate into classrooms. Thus, when teachers encounter contemporary content, they have three main choices: 1) to align new content to standards and teach it in full; 2) to follow district-mandated syllabi and pacing and ignore novel curricula; or 3) to scatter new materials throughout the mandated curriculum, hoping to engage students along the way.

To address the question of how teachers effectively implement curricula related to contemporary biology topics, given these constraints, we employed the modified fidelity framework (Perez *et al.*, 2016) as a lens to understand teachers' curricular adaptations following a neuroscience PD. Neuroscience served as an interesting, contemporary biology content area with limited links to national and state standards to examine teacher implementation strategies following PD.

METHODS

Context

High school science teachers from one urban and one suburban school district attended a two-week summer inquiry-based neuroscience PD workshop. The workshop, conceived and developed in conjunction with district-level science supervisors, was designed to teach key neuroscience ideas through inquiry-based instructional strategies appropriate for high school classrooms (MacNabb *et al.*, 2006; Dubinsky *et al.*, 2013). A sample workshop schedule has been previously published (Schwartz *et al.*, 2019). The districts were interested in the demonstrated potential of neuroscience instruction to impact student learning and in providing deep content knowledge to their teachers in a rapidly changing field (Blackwell *et al.*, 2007; Yeager *et al.*, 2019). PD goals were to increase teacher knowledge of two big ideas in neuroscience, synaptic plasticity and structure–function rela-

tionships, and inquiry processes sufficiently for them to be able to transfer that knowledge to their students (Supplemental Table 1). District supervisors helped recruit participating teachers and also attended the summer PD. Teachers were compensated for attending the workshops at union contract rates and received additional funding for classroom supplies to implement curricular activities modeled in the PD.

Guidelines for best practices in PD were followed (Desimone, 2011; Wilson, 2013; Supplemental Table 2). Effective PD provides teachers with adequate time to learn, practice, and reflect upon new learning. Research shows that PD interventions of at least 80 hours are necessary to promote positive changes in teachers' inquiry-based practices (Supovitz and Turner, 2000) and PD programs in the 30–100 hour range have larger impacts on student achievement than PD programs of 15 hours or fewer (Yoon *et al.*, 2007). Research also shows that PD should focus on content and actively engage teachers as learners in model science lessons (e.g., Kleickmann *et al.*, 2016; Lotter *et al.*, 2016) and authentic research opportunities (Enderle *et al.*, 2014; Yang *et al.*, 2020). Within BrainU, teachers learned neuroscience in the manner they were expected to use in their teaching. Within each activity, teachers communicated scientific ideas using the argumentation format of making a claim, supporting it with evidence, and providing reasoning to link the two. Each lesson was followed by discussion of the modeled pedagogical strategies and how they might be used in the classroom. These central PD activities within BrainU supported the transfer of PD content and pedagogy by using reform-based curricular materials. Research shows that the inclusion of curricular materials within PD improves both teacher and student outcomes (e.g., Lynch *et al.*, 2019; Roth *et al.*, 2019; Kowalski *et al.*, 2020).

Active learning also incorporates opportunities for teachers to implement and reflect on their implementation of new pedagogical approaches within their own classrooms (e.g., Roth *et al.*, 2019). During the workshop, teachers wrote action plans for how they were going to integrate the neuroscience lessons into the following year's curricula. All instructional decisions were made by the classroom teachers in both districts in accordance with state standards and district policies, paying attention to the importance of coherence between PD and schools (Desimone, 2009). Teachers knew their students much better than the PD providers and were encouraged to adapt lessons to their own classroom needs and district expectations. Common district-wide assessments were in place in both districts that guided all instructional decisions. During the academic year, program staff interacted with teachers to provide content, classroom support, and follow-up and to observe implementation.

Neuroscience concepts emphasized and included specialization of cellular structure, structure and function relationships, and how the nervous system commands and controls mental and physical functions with regulation and homeostatic feedback. Most importantly, the workshop emphasized the neurobiological basis of learning and memory through synaptic plasticity. Lessons and activities included in the PD can be found at www.brainu.org. Each lesson, as well as PD content and assessment, contained mapping to appropriate state and national standards (Supplemental Table 1). The most popular lesson, Altered Reality (AR), illustrated brain adaptability to novel sensory input, or “plasticity” in neuroscientific terms. In guided

TABLE 1. Student demographics^a

Student group	Participant		Control		All	
	N	%	N	%	N	%
Non-white ethnicity	597	54.7	191	43.3	788	48.6
ELL	282	25.8	76	17.2	358	23.4
FRL	574	52.6	173	39.2	747	48.7
SPED	59	5.4	18	4.1	77	5.0

^aCategories of students: ELL, English language learners; FRL, receiving free and reduced lunch; SPED, receiving special education. Students might be included in more than one row in this table.

inquiry, students investigate learning to toss beanbags at a target while wearing prism goggles. Students collect, interpret, and communicate their data, explanations, and conclusions. An open-inquiry extension provides students with an opportunity to design additional experiments.

Research Design

A concurrent mixed-methods research design was employed to meet the pragmatic and contextual demands of this study (Tashakkori *et al.*, 2015). Qualitative analysis explored teachers' implementation of neuroscience lessons to understand the nature of instruction following the PD. The qualitative analysis revealed implementation strategies that were subsequently used as categories in the quantitative analysis. The quantitative analysis examined student performance on a content test administered to students in participants' and control teachers' classrooms before and after neuroscience instruction.

Participants

Partner districts were located in the midwestern region of the United States. The urban district had a diverse student population, serving more 37,000 students with the following demographics: 14% Hispanic, 24% White, 30% Asian or Pacific Islander, 30% Black, and 2% American Indian or Native Alaskan. English language learners made up 18% of the student population. More than 70 languages were spoken, although only English, Hmong, Spanish, and Somali were used for school communication. Within this student population, 72% were eligible for free or reduced lunch. The suburban district served more than 39,000 students with the following demographics: 4% Hispanic, 78% White, 6% Asian or Pacific Islander, 10% Black, and 1% American Indian or Native Alaskan. English language learners made up 6% of the student population, and 30% of students were eligible for free or reduced lunch.

All teachers, supervisors, and student participants consented or assented, as appropriate, to participate in this institutional review board–approved study (protocol no. 0905S66761).

Teachers. The workshop was available to secondary school science teachers in these districts, promoting collective participation, which has been shown to have a larger impact on student learning than PD with single teachers from multiple schools and districts (Lynch *et al.*, 2019). In the suburban district, 12 teachers with from 1 to 34 years of experience participated, evenly distributed from among three high schools and an alternative learning center (Supplemental Table 3). In the urban district, 13 teachers with from 3 to 31 years of experience participated, evenly distributed among five high schools and an alternative learning center. Twenty of these teachers conducted pre- and

posttesting of their students and represented the study sample for exploring the effect of the PD on student learning compared with 11 control teachers from the same districts who taught neuroscience but had not participated in the PD. Control teachers had the opportunity to attend neuroscience PD in subsequent years. Some control teachers had been teaching neuroscience for many years and felt they did not need additional PD. Baseline observation data showed that control and treatment teachers were equivalent in terms of classroom practices.

Students. In total, pre- and posttest data were collected from 1122 students in classes with teachers who had participated in the neuroscience PD and 450 students in nonparticipant control teacher classes. Of the 1572 students with pre- and posttest data, 39 students were missing demographic information. These 39 students were not included in the hierarchical linear modeling (HLM) analysis. Information about the demographics of the students included in the study can be found in Table 1.

Data Collection

Teachers. A number of sources contributed to developing an understanding and classification of workshop participants' implementation strategies. These included teachers' initial written action plans developed at the end of the summer PD, formal classroom observations, teacher end-of-year reflections, and program staff field notes from observations and interactions with teachers during in-district follow-up sessions. Action plans included specific student learning goals, implementation strategies, assessment strategies, a list of activities to be implemented, and a set of personal goals for the teacher. At least twice during the school year following the PD, teachers were asked to complete a table listing which lessons they implemented and when.

Observations followed a written protocol that included a short pre-observation interview, detailed open field notes, and detailed and summative descriptions of what occurred (Roehrig *et al.*, 2012). Two external observers and four authors (C.L.E., G.H.R., M.C.H., R.A.H.) were trained on the observation protocol with prerecorded class videos, as previously described (Roehrig *et al.*, 2012). Training continued until interrater reliability reached or exceeded 90%. The majority of observations were performed by the external reviewers who were blinded to the PD participation, with the author-observers filling in during scheduling conflicts. Observation scheduling was always at the convenience of the teacher. Both participant and control teachers were observed at least once during the academic year following the workshop while implementing their neuroscience and non-neuroscience lessons. Non-neuroscience lesson observations were included to accommodate control teachers and make the participant status less obvious to the external

observers. Altogether neuroscience and non-neuroscience lessons were observed for 13 and 18 teachers who attended PD and four and 10 control teachers, respectively. Observers maintained open field notes from the classroom observations. Notes from a follow-up workshop, a year after the initial PD, in which teachers reflected upon the successes of their implementation strategies, were also used to understand the nature of their implementation. These data sources were used to triangulate and categorize teacher's implementation of neuroscience following the PD (see *Qualitative Determination of Implementation Strategy*).

Students. The student outcome variable in this study was the sum of a forced-choice neuroscience content knowledge test that was administered before and after teachers' neuroscience instruction. The neuroscience content test was written jointly by program staff and district science supervisors as recommended (Bass *et al.*, 2016). Research shows that the ability to detect the impact of PD on student outcomes is improved when student assessments are developed by the researcher to be aligned with the PD goals, as measures such as standardized state tests are too distal to detect impacts of PD (Lynch *et al.*, 2019). The student knowledge test was not meant to be a comprehensive assessment to gauge complete mastery of neuroscience content covered in the PD. Because teachers had the freedom to implement lessons of their choice, the assessment had to focus on the big ideas that were covered in multiple different lessons (Supplemental Table 1). The district administrators also did not want this assessment to take up more than 20 minutes of class time, so as not to compromise overall instructional time. Thus, a relatively short test was designed to sample student knowledge. As such, it should be viewed as a measure for the relative effectiveness of the teaching, not as an absolute measure of complete student knowledge. Teachers had access to their own students' test scores, to use formatively or summatively as they chose. The student assessment was used in the year before the PD to familiarize all teachers, including controls, with its content and the process of distributing and collecting the tests.

Test items were written to cover content and align with state and national standards for brain structure–function relationships, synaptic plasticity, and scientific processes (Supplemental Table 1). Some test items had been used in previous studies and were evaluated there (MacNabb *et al.*, 2006; Roehrig *et al.*, 2012). Item difficulty varied to prevent ceiling effects. Item analyses of the neuroscience content test found that five of the items did not have sufficient discrimination (i.e., they were very difficult and most students did not answer them correctly) and were dropped from further analyses on both pre- and posttests (see Supplemental Material, Student Assessment and following, for details on the student assessment). A reliability study of the instrument on the complete set of students' posttest scores from participant teachers indicated that the instrument was reliable ($\alpha = 0.65$, SEM = 1.89; Supplemental Material).

Data Analysis

Research Question 1. Effect of Professional Development
Quantitative Analysis. Both descriptive and hierarchical statistics were employed. Conventional statistical analysis, including *t* tests and one-way analyses of variance were used for prelimi-

nary analysis as indicated in the text (Microsoft Excel). To study the effects of teacher PD on subsequent student achievement, we must take into consideration the nesting of students within classrooms. In real-world educational settings, student placement within classes is not random but rather is based upon the school's scheduling limitations. Students within a classroom are likely more similar to one another than they are to a random sample from a larger population. Such clustering violates the assumption of independence, making it unlikely that student ability will be equally distributed across all classes. Similarly, differences in teachers due to their years of experience teaching or whether they were teaching advanced or regular biology may contribute to student performance. To capture and attribute variance to these intersecting student- and teacher-level characteristics, we used HLM (Raudenbush, 1988; Bryk and Raudenbush, 1992). In HLM, students were nested within their classrooms, represented by their teachers, reflecting the nonrandom scheduling described earlier, among other factors. The nesting prevents aggregation bias and permits distinguishing variance attributable to both student and teacher levels. Variations in student outcomes would be expected to be due to differences in teachers and their implementations of the curriculum as well as to student characteristics. A single-level linear regression model would not be expected to capture these dependent variations (Raudenbush, 1988; Van Dusen and Nissen, 2019). The use of an HLM model prevents type 1 errors (falsely finding significance) and incorrect inferences (Young *et al.*, 1996; Raudenbush, 1988; Van Dusen and Nissen, 2019).

Students comprised the first level, *i*; teachers comprised the second level, *j*. The number of participating teachers in some schools was too small to permit comparisons at the school level. Partnering with two different large school districts could have produced effects attributable to the district level as well. However, such comparisons were not the focus of the research questions or in the interests of the partner districts. The study did not capture a large enough *N* to model every possible teacher or structural variable. Social and economic differences between the districts were represented in the student-level demographic variables. Differences among the teachers, who all took the PD together, would be captured in the observational-level teacher variables. The dependent variable was student performance on a neuroscience knowledge test administered in a pre–post manner in all classes.

The first HLM analysis determined how teacher PD itself influenced overall student neuroscience learning. HLM was deemed the most appropriate analysis to account for variations and dependencies occurring at both student and teacher levels while acknowledging that students were nested within teachers' classrooms (Raudenbush, 1988). Whether a teacher had participated in the PD (control vs. treatment teachers) was coded as a dichotomous variable. The overall analysis of student outcomes was performed using HLM v. 6.0 (Scientific Software International).

A series of pilot HLM models determined that covariation among teacher variables was insignificant and marginally significant among student variables, although this was not of practical concern given the large sample size. Allowing slopes to vary did not improve the statistical model, so they remained fixed. Adding teacher-level variables, such as years of teaching experience or teaching specialty, to the final model continued to

explain variance, but resulted in identification of single teachers. To prevent this, we did not use additional teacher variables. The final model was:

$$\begin{aligned} \text{POSTTEST}_{ij} = & \gamma_{00} + \gamma_{01} * \text{PARTICIPANT}_j + \gamma_{10} * \text{PRETEST}_{ij} \\ & + \gamma_{20} * \text{MINORITY}_{ij} + \gamma_{30} * \text{FRL}_{ij} + \gamma_{40} * \text{SPED}_{ij} \\ & + \gamma_{50} * \text{ELL}_{ij} + \gamma_{60} * \text{FEMALE}_{ij} + u_{0j} + r_{ij} \end{aligned}$$

where PRETEST and POSTTEST represent scores for student i in teacher j 's class. PARTICIPANT represents whether the teacher had participated in the PD or not. FRL, SPED, and ELL represent student status on free and reduced priced lunch (FRL), special education (SPED), and English language learner (ELL). MINORITY represents whether the student was American Indian, Asian, Pacific Islander, Hispanic, or Black or not and FEMALE represents whether the student was female or not. The pretest variable was group-mean centered (on the mean for each teacher's classrooms) to afford comparisons between teachers according to their treatment. All other independent variables were dichotomous and were not centered. γ_{00} represents the grand mean performance, after ascribing variance to the appropriate variables. u_{0j} represents the residual variance due to teacher-level groupings of students. r_{ij} represents the residual variance attributable to students.

Research Question 2: Effect of Curricular Implementation. Research question 2 employed a sequential mixed-methods approach (Tashakkori *et al.*, 2015). First, the qualitative data were used to determine types of curricular implementation resulting from teacher decisions. Second, categories arising from the first of two common decisions were used within a second HLM analysis to determine the effect of implementation type on student outcomes.

Qualitative Determination of Implementation Strategy. A collaborative and iterative process was followed to extract meaningful themes from the variety of observations and teacher artifacts (Miles and Huberman, 1994). All data sources listed earlier were examined for ideas related to implementation strategies. The timing and inclusion of specific lessons or categories of lessons were examined for commonalities regarding content, pedagogy, and embrace or rejection of inquiry. Observation data and field notes were combed for evidence of larger strategies that teachers might be invoking, like scaffolding or specific sequencing with respect to content or inquiry. An initial set of codes (M.C.H., C.L.E., & R.A.H.) was developed and applied to all artifacts by three author-observers. This set of codes included various decisions teachers made either in their action plans or in their delivery. Iterative discussions about the codes led to the development of descriptive themes. This iterative process continued until full agreement was reached and a preliminary analysis was written. Subsequently a different set of three authors (C.L.E., G.H.R., & J.M.D.), including one from the original analysis group, revisited the artifacts to verify the coding and thematic clustering. Once the overarching themes of unit versus sprinkling emerged (see *Sprinkling* section), teachers were categorized as following one of these strategies based upon their periodic responses listing lessons and dates delivered. The

implementation strategies uncovered in this manner were subsequently used in the quantitative analysis.

Quantitative Analysis of Implementation Strategy. The second HLM analyses focused on how teacher implementation strategies influenced overall student neuroscience learning. Teachers did not alter their implementation strategies between multiple classes in a day, so the placement of students within particular class periods was not considered as a variable. Additionally, the classroom-level implementation strategies were coded dichotomously in separate variables for each strategy. All variables were represented as being present or absent, that is, having values of 1 or 0, respectively. Because these observationally derived characteristics were not known a priori, no assumptions were made concerning relationships among them. This model contained the implementation strategies of having a neuroscience unit or sprinkling, as determined in the qualitative analysis (see definitions under discussion of research question 2 in the *Results* section). Demographic variables that were not found to be significant predictors in the previous model were tested for significance in the implementation strategy model, but again did not explain significant variance. Because the same demographic variables were found to be significant in both models, the final HLM model for determining the effect of implementation strategy was:

$$\begin{aligned} \text{POSTTEST}_{ij} = & \gamma_{00} + \gamma_{01} * \text{NEURO UNIT}_j + \gamma_{02} * \text{SPRINKLING}_j \\ & + \gamma_{10} * \text{PRETEST}_{ij} + \gamma_{20} * \text{FRL}_{ij} + \gamma_{30} * \text{SPED}_{ij} \\ & + u_{0j} + r_{ij} \end{aligned}$$

where PRETEST and POSTTEST represent scores for student i in teacher j 's class. NEURO UNIT and SPRINKLING designate the teaching strategies implemented. FRL and SPED represent student status on free and reduced priced lunch (FRL) and special education (SPED). As with the previous model, the pretest variable was group-mean centered. All other variables were dummy coded and were not centered.

RESULTS

Research Question 1. Effect of Professional Development

To examine the overall effect of the PD on student learning, we analyzed student neuroscience tests for the treatment and control teachers. In general, there was an overall increase in performance seen from the pretest to the posttest for students across all student groups (Table 2). An independent t test comparing the mean posttest scores of students in classrooms with teachers who attended the PD ($M = 6.66$, $SD = 2.79$) with scores of students in control teachers' classrooms ($M = 5.58$, $SD = 2.48$) found that the treatment group had significantly higher posttest scores ($t(1570) = 7.17$, $p < 0.001$, $d = 0.41$; Supplemental Table 4). However, this aggregate analysis does not consider variation among students that must be considered. HLM statistics permit attributing variation to both student and teacher inherent variables as well as the PD and implementation variables under examination.

The HLM model, which controlled for student demographics, student pretest scores, and teacher effects, also found a significant improvement in student posttest scores when the

TABLE 2. Pre- and posttest student scores by demographic group^a

Student group	Percentage of sample	Pretest		Posttest	
		M	SD	M	SD
All students	100.0	5.43	2.27	6.41	2.77
Minority	50.4	5.11	2.16	6.14	2.50
ELL	22.7	4.88	2.03	5.97	2.39
FRL	47.3	5.06	2.17	6.03	2.52
SPED	4.8	4.54	2.49	5.70	2.82

^aScores represent the number of correct responses on the 13-question pretest and posttest (mean \pm SD). Categories of students: minority, any non-White student; ELL, English language learners; FRL, receiving free and reduced lunch; SPED, receiving special education.

teacher had participated in the PD (Table 3). The coefficient column in Table 3 indicates the projected amount of student performance points attributable to the indicated variables. Only the student pretest scores, whether the student was receiving free or reduced lunch, and whether the student was receiving special education services were found to be significant predictors in the model. Students' minority status was not a predictor of student performance; students of all ethnic, cultural, and racial backgrounds learned neuroscience with equal gains from pre- to posttest. In subsequent models, minority, English, and gender status were dropped from consideration.

Research Question 2. Effect of Curricular Implementation

Overall, implementation strategies involved two independent decisions. The first teacher decision was whether to enact a complete neuroscience unit or to scatter neuroscience lessons throughout existing biology course units (coded as "unit" or "sprinkling"). Unit implementation generally involved devoting 1 to 2 weeks to contiguous neuroscience content. Integrating individual neuroscience lessons into other biology units spread throughout the course was termed "sprinkling" by the teachers. The second teacher decision was how and when to implement the popular Altered Reality inquiry lesson that illustrated synaptic plasticity. Implementation at the beginning of the school year often emphasized inquiry processes more than the plasticity message. Beyond Altered Reality, teachers did not execute common clusters of lessons (Supplemental Table 1). Other ideas that arose in the

coding were not sufficiently universal to be able to categorize teachers' implementation strategies for use in the subsequent HLM analysis.

Implementation Strategies

First Decision. Following the workshop, the teachers' first decision focused on how to integrate the neuroscience content into their biology courses. The action plans they wrote during the workshop facilitated this decision and served as outlines that included lesson choices and sequencing judgments. Classroom observations, field notes from teacher interactions, and lists of lessons and dates corroborated implementation strategies and revealed two broad strategies: implementing a stand-alone neuroscience unit (12 teachers) versus sprinkling neuroscience lessons into other biology units (eight teachers). In making this decision, teachers used knowledge gained from the PD and associated curricular materials, their experiences using these materials during the PD, and their personal knowledge of the mandated biology curriculum.

Sprinkling. Teachers were clear that sprinkling was a conscious decision; as one stated, "We will strategically embed neuroscience into a variety of appropriate curricular areas." This decision was often guided by available time, as one teacher stated, "Because class time and preparation time are serious constraints, the current plan involves rotating some of the activities in and out of the curriculum." District emphasis on teaching to standards also drove this decision. One action plan

TABLE 3. Results from the HLM model on the effects of teacher PD with student demographics^a

Fixed Effect	Coefficient	SE	t ratio	p
Control classroom mean achievement, γ_{00}	5.70	0.29	—	
Participant teacher effect, γ_{01}	0.97	0.41	2.38	*
Pretest slope, γ_{10}	0.37	0.02	14.93	***
Minority, γ_{20}	-0.12	0.23	-0.54	
FRL, γ_{30}	-0.36	0.15	-2.33	*
SPED, γ_{40}	-0.57	0.24	-2.39	*
ELL, γ_{50}	-0.24	0.18	-1.30	
Female, γ_{60}	0.22	0.12	1.84	
Random effect	Variance	df	X ²	p value
Teacher mean, u_{0j}	1.39	29	599.93	<0.001
Student effect, r_{ij}	4.75			

^aIn the HLM model, the coefficients represent the calculated value of the 1) average response after correcting for the model variables, γ_{00} ; 2) additional increments in response values expected or attributable to the teacher participating in the PD, γ_{01} ; student pretest values, γ_{10} ; student demographics, γ_{20} , γ_{30} , γ_{40} , γ_{50} , γ_{60} . Positive coefficients indicate a higher score; negative coefficients indicate a lower score. * $p < 0.05$; *** $p < 0.001$. Robust standard errors have been used.

explicitly stated, “Goal: Integrating new curriculum into existing curriculum/standards.”

Neuroscience information was typically added to existing units on cells, proteins, and human organ systems. For example, “I plan on using the neuron as an area of focus during cells. I plan on using the nervous system and sheep brain as an area of focus during organ systems.” This implementation approach suggested teachers recognized how the diversity of neuronal form and function made it an example of a complex cell that could illustrate many cellular functions. One teacher used neuroscience as a strategy to make her biology instruction more holistic by emphasizing the

interrelationships of the concepts of biology as they apply to the brain. Neurons are brain cells. They can receive and carry electrical and chemical impulses. Cells have a lipid bilayer that is semi-permeable with a fluid mosaic of proteins embedded, either spanning the membrane or riding on the periphery of it. I knew those two topics and taught them to students. However, it was through the discussions with neuroscience between myself and my colleagues ... that I connected the two topics and was able to gain a holistic view of the cell membrane as it related to sending signals in the form of ions and neurochemicals.

Behaviors served by neuronal circuits and connectivity were used by teachers to illustrate the crosscutting concept of systems function. One action plan stated, “The brain is a system that can be investigated; models can be made and refined, and subsystems of the brain can be identified by students (through experimentation with behavior of model organisms such as *C. elegans* or through dissection of sheep brains for example) and connections can be identified.” Neuroscience was also integrated into units on homeostasis, diversity of organisms, animal behavior, and scientific inquiry.

Unit. Intensive neuroscience units reflected teachers’ desires to engage their students in a learning process emulating the one they experienced at the workshop. Specifically, teachers desired to provide multiple active-learning experiences for their students. One teacher justified a stand-alone unit in a reflection on his own neuroscience learning:

In this short period of time I had a much more profound understanding of a neuron and its function. I believe this was the result of not just one of the activities [the PD] developed around this concept, but rather an accumulation of all the diverse experiences I received around this topic. I would not have had the exposure to the technical aspects of neural transmission without [the] lecture, and I would not have a use for this knowledge if it were not for the neuron food model. In addition, without the lab on beady neurons I would not understand how neurons make circuits. All three activities related to the same topic, but all three activities illustrated different parts of the whole. It is this aspect of [the PD] that has impressed upon me the most as an educator, the ability to take concepts and illustrate them through diverse and numerous ways.

Units were also enacted when teachers stated broader goals that included scientific processes in addition to content learning goals. For example, one action plan provided the following details:

I intend to create a more hands-on, inquiry-based learning environment. It is also my goal that students better understand the structure and function of neurons and the nervous system, that they increase their capacity to reason and think individually, and that they improve their ability to plan and carry out self-guided scientific investigations. Goal: Improvement of student abilities in reasoning, critical thinking, and the scientific method.

Second Decision. The second pedagogical decision emerging from the qualitative analysis, independent of the first, was the pedagogical emphasis of the most popular and hence observed neuroscience lesson, *Altered Reality*. The lesson could be implemented as part of the neuroscience content emphasizing synaptic plasticity as the basis for learning and memory or it could be implemented to emphasize general scientific practices or both. Teachers recognized the engaging nature and dual learning goals inherent in the AR lesson, with many teachers explicitly stating their goal was to teach scientific processes and raise student awareness of the biological basis of learning. As one teacher said, they were inspired “to illustrate the scientific method and to implant an open mindedness about learning and plasticity of the brain.” The desire to engage students with scientific processes motivated eight teachers to use the popular AR lesson at the beginning of the school year to introduce students to scientific practices in preparation for further future experimentation. Other teachers used the AR lesson within a unit later in the year (eight teachers) or did not enact it at all (four teachers). Only half of the AR lessons were observed, as we did not know beforehand that the pedagogical approach to this lesson was going to emerge as a critical implementation decision.

Emphasis on Plasticity. The plasticity message was front and center alongside inquiry processes in two observed classrooms. For example, one teacher guided students through the inquiry process with the entire class agreeing on a single throwing procedure on day 1 and gathering data. This was followed up on day 2 with an in-depth discussion interpreting the data through the lens of synaptic plasticity and learning before permitting the class to design their own goggle experiments. The observer noted, “The lesson was framed around a discussion of learning and the idea that we can change our brains based on what experiences we embrace with our attention and cognition.”

Emphasis on Scientific Processes. Six observations of teachers’ first AR lessons of the year revealed a focus on an open-inquiry approach to the AR experiment. The teachers skipped the initial guided-inquiry experiment with a plasticity message in favor of emphasizing student-designed experiments to model scientific processes in the recommended extension part of the lesson. In these six classes, groups of students brainstormed and picked a question of interest for investigation using the prism goggles. Typical student questions included: “What is the effect on the accuracy of the throw with the goggles on?” “Will throws get better over time as you get used to the goggles?” “How do reality goggles affect balance?” Students were highly engaged throughout the AR activity and demonstrated a lot of curiosity and creativity in designing throwing and walking challenges wearing the goggles. In all of the observations, students

were expected to present their results to the whole class, either orally or as posters accompanied by a “gallery walk,” where students walked around and commented on other groups’ posters. Teacher questions during group presentations focused on making sure that students understood variables, appropriate graphical representations of data, and modifications to procedures for a hypothetical repeat data collection.

Observations on the AR lesson suggest that an open inquiry alone approach was ineffective at developing the idea that synapses change, as modeled during the PD. Teachers focused on scientific practices such as how to write a researchable question, how to control variables, and how to visualize and graph data. Students explored questions such as differences in throwing ability by gender or handedness (right vs. left), which did not lend themselves to conclusions about how the brain adapts to the AR goggles. To develop the concept of brain plasticity, it is critical that students explore how their ability to hit a target improves over repeated attempts as the brain adapts to the visual distortion created by the prism goggles. In opening up the inquiry to any student-generated questions, the idea of brain plasticity was lost in these observed lessons, as students explored questions about girls versus boys, right-handed throws versus left-handed throws, and so on. Opening up the inquiry experience, rather than maintaining a guided-inquiry approach as modeled in the PD, teachers were successful in meeting their goal of introducing inquiry skills but appeared unsuccessful in reaching their neuroscience content learning goal. Unfortunately, as the different approaches to the AR lesson emerged from the analysis and not every use of this lesson was observed, the effect of this second implementation decision on student performance could not be examined.

Effect of the Unit/Sprinkling Decision on Student Learning. To determine the effect of the first individual PD teacher implementation decision, we assigned each teacher post hoc to a category of either unit or sprinkling, as determined in the qualitative analysis by reviewing the reported lists of lessons delivered on certain dates. No significant difference was found between the number of neuroscience lessons covered by teachers who sprinkled or enacted units (8.1 ± 6.6 lessons/teacher who sprinkled, 6.9 ± 5.5 lessons/teacher using units, $p = 0.66$, $d = 0.19$, two-tailed t test). This similarity extends to a breakdown of lessons taught by the three assessment domains of

interest (Supplemental Table 1). Descriptive-level statistics suggested that, when teachers taught a full unit, students performed better. However, variations in student pretest values limited the interpretation of these calculations (Supplemental Table 4). Thus, HLM analysis was again performed to account for student variability. The unit and sprinkling teacher-level categories were added to the HLM model in place of the PARTICIPANT teacher-level variable. These were added as separate variables so that the impact of each strategy could be evaluated independently compared with controls. When controlling for teachers’ implementation strategy and student pretest scores, only free or reduced lunch (FRL) and special education (SPED) were again found to be significant student-level predictors of posttest scores. The coefficient column in Table 4 indicates the projected amount of student performance points attributable to the indicated variables. After controlling for student demographics, we found that PD-trained teachers who used an intensive unit had an average predicted student posttest score that was significantly higher than the projected grand mean ($SE = 0.72$, $t = 2.76$, $p < 0.05$), whereas teachers who implemented the sprinkling strategy did not have significantly higher student posttest scores ($SE = 0.49$, $t = 1.30$, ns).

To determine whether the implementation strategy had a differential effect upon any of the three domains probed in the student assessment, we separately analyzed student performance on the sets of questions relating to synaptic plasticity, nervous system structure–function, and inquiry processes. Because the range of scores on the test subdivisions were limited, HLM could not be used. Student pretest to posttest gains or change scores were calculated for the entire test and for each subdivision (Figure 1; for effect sizes, see Supplemental Table 5). In this analysis, the gains in learning on the entire test were greater for students in unit classrooms compared with those in sprinkling and control classrooms (Figure 1A). Student learning gains were also greater in classrooms of PD teachers who sprinkled compared with control teachers (Figure 1A). Both were expected from the HLM analysis (Table 3). For the plasticity and structure–function domains, change scores in unit classrooms significantly exceeded those in control classrooms (Figure 1B and C). This suggests that, even though some teachers were not observed addressing plasticity during their AR lessons, this concept was integrated into the unit during lessons following AR. For the inquiry domain, PD classroom students, regardless of

TABLE 4. Results from the HLM model of teacher implementation with student demographics^a

Fixed effect	Coefficient	SE	<i>t</i> ratio	<i>p</i>
Classroom mean achievement, γ_{00}	6.88	0.36	—	
Intense neuroscience unit effect, γ_{01}	1.98	0.72	2.76	*
Sprinkling neuroscience effect, γ_{02}	0.64	0.49	1.30	
Pretest slope, γ_{10}	0.34	0.03	12.34	***
FRL, γ_{20}	−0.56	0.22	−2.53	*
SPED, γ_{30}	−0.79	0.28	−2.88	*
Random effect	Variance	<i>df</i>	X^2	<i>p</i> value
Teacher mean, u_{0j}	1.72	29	408.85	<0.001
Student effect, r_{ij}	6.22			

^aIn the HLM model, the coefficients represent the calculated value of the 1) average response after correcting for the model variables, γ_{00} ; 2) additional increments in response values expected or attributable to the teacher implementation strategies, γ_{01} , γ_{02} ; student pretest values, γ_{10} ; student demographics, γ_{20} , γ_{30} . Positive coefficients indicate a higher score; negative coefficients indicate a lower score. * $p < 0.05$; *** $p < 0.001$. Robust standard errors have been used.

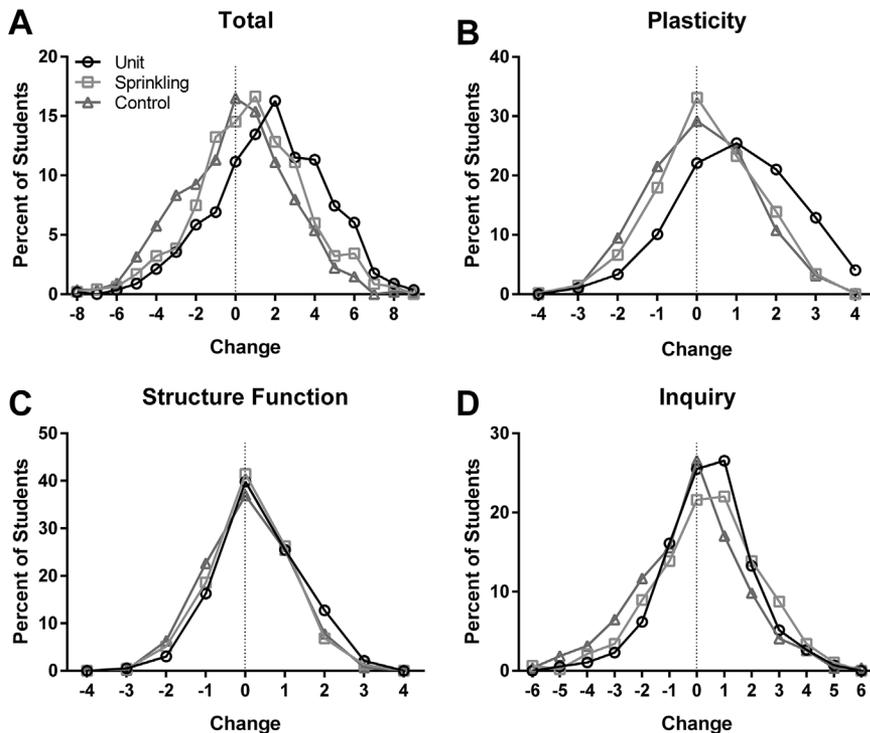


FIGURE 1. Histogram contours of change scores (posttest minus pretest) for the entire test (A) and the plasticity (B), structure–function (C), and inquiry (D) subdivisions are plotted as connected data points. Each data point represents the height of a histogram bar with a bin size of 1 point. Lines connecting the data points represent the envelope of the histogram for students in classrooms of PD teachers who implemented units (open circles, black), PD teachers who implemented sprinkling (open squares, light gray), and control teachers (open triangles, medium gray). Thus, each graph depicts three overlapping histogram contours for the different classroom implementations. Shifts to the right indicate greater changes in scores or more student learning. For statistical comparisons for the data in each graph, see Supplemental Table 5. Total student $N = 1572$.

teacher unit or sprinkling strategies (which were equivalent), outperformed control classroom students (Figure 1D). Thus the PD improved teachers' ability to teach inquiry irrespective of implementation strategy. Unit implementation produced greater student learning of both plasticity and structure–function content than either sprinkling implementation or control.

DISCUSSION

Overall, the neuroscience PD resulted in greater student learning of neuroscience content compared with students of teachers who did not attend the PD, which is consistent with prior research (Lee *et al.*, 2005; Akerson and Hanuscin, 2007; McNeill and Krajcik, 2008). Darling-Hammond and colleagues (2017) noted the lack of rigorous research using quasi-experimental and randomized control tests providing strong empirical support for the impact of PD on student outcomes. Our quasi-experimental design used a rigorous HLM analysis, adding to the limited data supporting the impact of PD on student learning in science classrooms. Of note, this is only the second study providing empirical support for the impact of PD on high school science learning (Greenleaf *et al.*, 2011), with the majority of prior research using quasi-experimental or randomized control designs having taken place in elementary classrooms (Marek

and Methven, 1991; Buczynski and Hansen, 2010; Johnson and Fargo, 2010, 2014; Roth *et al.*, 2011; Heller *et al.*, 2012; Kleickmann *et al.*, 2016; Taylor *et al.*, 2017).

Central to the design of our PD was the availability of quality curricular materials to support implementation of neuroscience lessons in high school classrooms. Our findings support the current literature (Buczynski and Hansen, 2010; Johnson and Fargo, 2010, 2014; Heller *et al.*, 2012; Kleickmann *et al.*, 2016) that calls for involving teachers in active learning as students in model lessons using the classroom curricular materials. Our prior work showed that this has an impact on teachers' content knowledge and classroom practices, and this study confirms the positive impact on student learning (MacNabb *et al.*, 2006; Roehrig *et al.*, 2012; Schwartz *et al.*, 2019). Our PD used high levels of scaffolding (Doppelt *et al.*, 2009), both by a neuroscience expert who was able to support the development of teachers' content knowledge through inquiry-based activities and a science education expert who provided discussions of pedagogy and models of inquiry-based teaching (Penuel *et al.*, 2011). In reviewing the lessons that teachers selected and sequenced for implementation, it was clear that they were intentional in maintaining inquiry-based teaching approaches even when modifications were made. Thus, like Penuel and colleagues (2011), we argue that helping teachers to develop

a strong understanding of the content and providing strong learning models through curricula is important in promoting curricular implementations that align with the intentions of the PD and written curriculum. Teachers' implementation ranged from moderate to high fidelity to the goals promoted within the PD to promote learning neuroscience through inquiry-based lessons. Even in cases in which curricular modifications were made during implementation, such as was described for the AR lessons, the PD teachers' neuroscience instruction showed stronger student learning than the control group. Critical to the field, this study extends understanding of the effects of PD on student learning by exploring the different ways in which teachers implement PD-provided curricula and how these curricular modifications impact student learning. While other research acknowledges that teachers will modify curricula to their teaching context (Davis and Varma, 2008; Johnson and Fargo, 2010; Penuel *et al.*, 2011; Hill and Erickson, 2019), our study specifically investigated how teachers' curricular modifications and implementation strategies for a contemporary biology topic introduced during PD effected student learning. While we did not explore internal factors, such as teacher beliefs, that might impact implementation decisions (Roehrig and Garrow, 2007; Buczynski and Hansen, 2010), teachers shared external factors

that influenced the ways in which they decided to include neuroscience lessons into their high school biology curricula. Unlike other studies that report lack of materials and resources as the largest barrier to implementation (e.g., Buczynski and Hansen, 2010; Darling-Hammond *et al.*, 2017), our teachers were provided with the necessary classroom materials. However, similar to teacher experiences reported in the literature (Powell and Anderson, 2002; Roehrig and Garrow, 2007; Buczynski and Hansen, 2010; Darling-Hammond *et al.*, 2017), our teachers struggled with the extensive list of biology standards in the state science standards, fitting neuroscience into their district pacing guides and aligning with common district assessments.

Teachers did not always teach neuroscience in the exact manner or sequence in which they had learned it during the PD. Unit implementation most closely mimicked the teachers' own learning during the PD, as the PD was concentrated into 2 weeks over the summer. Teachers learning goals most closely matched those they experienced during the PD. Teachers who sprinkled neuroscience lessons throughout the biology course had to adapt and apply neuroscience concepts as examples of other biology principles; in other words, their learning goals were not centered on neuroscience content. This adaptation may have diluted the neuroscience messages, given that student content learning was significantly higher in unit implementation compared with sprinkling. This contrast is important not only for immediate learning, but also future learning, as spending more time going deeper into a high school biology unit resulted in higher student performance in subsequent college courses (Schwartz *et al.*, 2008).

Concentrating the content into a unit may have facilitated the development of students' conceptual understanding, as they were able to build knowledge through a coherent sequencing of concepts and investigations supported through the curriculum and experienced by the teachers in the PD (Doppelt *et al.*, 2009; Buczynski and Hansen, 2010; Heller *et al.*, 2012; Kleickmann *et al.*, 2016). In contrast, sprinkling may not afford teachers the opportunity to enact a cohesive storyline to build the neuroscience concepts into an understandable and memorable message for students. Like Hanuscin and colleagues (2016), we argue that, by simply doing single, isolated activities, teachers will fail to construct a scaffolded learning sequence or storyline, thereby minimizing the development of science concepts. By interspersing the neuroscience lessons into units focused on other biology concepts, the important neuroscience concepts of plasticity and structure–function relationships were not learned as well (see Supplemental Table 5 for effect sizes). Interestingly, when looking at the student assessment items related to inquiry, the mode of implementation was not predictive of student learning. As seen in the AR lesson observations and in the comparable gains made on the inquiry portion of the assessment (Figure 1D), teachers were better able to implement the pedagogical aspects of the PD than the content-focused aspects. Teachers successfully translated their own learning of inquiry processes to their students without this being tied to specific content.

The variability of the student outcomes studied here serves as an example for how PD on other contemporary biology topics may influence student learning. If teachers try to touch on novel topics encountered in PD or integrate or fold them into other larger units, student content learning may be less than expected. For contemporary science topics, for which curricular resources

are scarce and alignment with standards is not direct, teachers' decisions become important for integrating these often motivating and exciting ideas into the prescribed curricula. Despite promotion of and participation in the neuroscience PD by district supervisors, many teachers in our study were reluctant to spend time on neuroscience outside the mandatory biology topics more strongly aligned with statewide and district testing (Powell and Anderson, 2002; Roehrig and Garrow, 2007; Buczynski and Hansen, 2010; Darling-Hammond *et al.*, 2017). Thus, integrating the contemporary content throughout the year was a popular teacher strategy aimed at satisfying the often-conflicting goals of teaching to standards and engaging students with cutting-edge biology topics as modeled in the PD. By choosing a sprinkling “compromise” to meet these divergent goals, teachers fulfilled the goal of increasing neuroscience learning to a lesser extent than teachers implementing a neuroscience unit.

Limitations

One caveat of the data reported here is that teachers controlled the timing of the student posttest. We requested end of the year administration, but as some teachers used the posttest for their own assessment, we could not control its timing. The assessment may have been administered immediately after unit delivery or at the end of the school year, depending upon teacher implementation strategies or available class time. Unit presentation may simply have consolidated the material closer in time to the knowledge test than for sprinkled classrooms. Either way, the conclusions that the PD positively affected inquiry-based teaching and student learning compared with control teachers is unaffected by this caveat.

Additionally, our observations were unable to capture everything that goes on in classroom implementation. We were rarely able to observe a classroom during several neuroscience lessons. Observations were not exhaustive, missing some teachers and many lessons. Thus, these data do not speak to how teachers connected ideas established: 1) during the beginning of the year enactment of the AR lesson to later lessons in either unit or sprinkling classrooms; or 2) between one sprinkled lesson to another. Furthermore, the broad focus on implementation of the entire curriculum limited the ability to capture fine-grained insights into any but the most popular AR lesson.

Students' socioeconomic status (SES) is documented as a strong predictor of achievement (e.g., Sirin, 2005; Baker and Johnston, 2010; Gustafsson *et al.*, 2018). In our analysis, SES is accounted for using the students' FRL status. We did not use other variables such as zip code or grade school attended, which also could have captured performance differences based upon circumstances beyond FRL status. It should also be noted that additional student-level variables, such as prior science achievement and science interest, could have impacted student performance and impacted the model (Young *et al.*, 1996); however, these factors were not the focus of the current project.

Future Research

We did not have sufficient follow-up discussions with teachers to fully probe their implementation decisions regarding individual lessons. The content versus process emphasis observed in the AR lesson implementation deserves further investigation. There is some suggestion that teachers who sprinkled did not necessarily emphasize take-home messages or regularly connect

the lesson content to larger ideas. If the observed sample of neuroscience lessons is representative, this absence of lesson summation could explain the decrease in student performance. More simply, teachers could have felt less obligation to cover the subject and therefore did not invest much of their own intellectual energy. The observed weaker effectiveness of the sprinkled lessons suggests future investigations of implementation following PD should plan to follow teachers more closely and include interviews to capture teacher justifications for their pedagogical choices.

CONCLUSION

In the current study, teachers were given the autonomy to implement the neuroscience lessons as they saw fit, exercising judgment regarding how to organize and adapt the entire neuroscience curriculum to their own biology scope, sequence, and students. While teachers embraced the lessons for the enthusiasm they generated among students (e.g., the AR lesson), overall student learning was greatest when the implementation was cohesive, contiguous in time, and delivered as a unit. When considering the challenge of balancing exciting contemporary biology against the standards-based content requirements, teachers' decisions matter.

Prior studies have uncovered similar misalignments between classroom enactments and the goals of curricular materials (Davis *et al.*, 2011). Observational studies on curricular implementation rarely go beyond examination of the relationships between teachers and resources to examine the effect of teachers' implementation decisions regarding these resources on student outcomes. Yet teachers' curricular decisions have consequences for student learning. Importantly, this study is the first to demonstrate how teachers' decisions can affect student performance. Among this teacher cohort, two critical implementation choices stood out: how to integrate novel content into their scope and sequence and how to use the AR lesson. The teachers' decisions to implement a neuroscience unit or sprinkling of neuroscience lessons into other biology units were critical, as student learning was significantly greater in classrooms where teachers implemented a neuroscience unit.

ACKNOWLEDGMENTS

The authors wish to thank all the participating teachers and administrators for their input, cooperation, and eagerness to learn neuroscience. In addition, Selcen Guzey, Michael Michlin, Ken Jeddelloh, Devarati Bhattacharya, and J. McClelland contributed to the organization and evaluation of the teacher workshops, data collection, and analysis. Funding was provided by National Institutes of Health SEPA R25 OD011131 and the University of Minnesota Medical School to J.M.D.

REFERENCES

- Akerson, V. L., & Hanuscin, D. L. (2007). Teaching nature of science through inquiry: Results of a 3-year professional development program. *Journal of Research in Science Teaching*, 44(5), 653–680.
- Baker, M., & Johnston, P. (2010). The impact of socioeconomic status on high stakes testing reexamined. *Journal of Instructional Psychology*, 37(3), 193–199.
- Bania, M., Chase, V., Roebuck, B., & O'Halloran, B. (2017). *Local adaptations of crime prevention programs: Finding the optimal balance between fidelity and fit key effective elements of crime prevention programming (2017–R019-A)*. Ottawa, ONT, Canada. Retrieved September 13, 2021, www.researchgate.net/publication/327118483_Local_Adaptations_of_Crime_Prevention_Programs_Finding_the_Optimal_Balance_Between_Fidelity_and_Fit_A_Literature_Review
- Bass, K. M., Drits-Esser, D., & Stark, L. A. (2016). A primer for developing measures of science content knowledge for small-scale research and instructional use. *CBE—Life Sciences Education*, 15(2), rm2.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246–263.
- Blanchard, L. S., Southerland, S. A., & Granger, E. M. (2009). No silver bullet for inquiry: Making sense of teacher change following an inquiry-based research experience for teachers. *Journal of Research in Science Teaching*, 93(2), 322–360.
- Blank, R. K., & de las Alas, N. (2009). *Effects of teacher professional development on gains in student achievement. How Meta Analysis Provides Scientific Evidence Useful to Education Leaders*. The Council of Chief State School Officers: Washington, DC. Retrieved September 13, 2021, from <https://eric.ed.gov/?id=ED544700>
- Brown, M. W. (2009). The teacher–tool relationship: Theorizing the design and use of curriculum materials. In Remillard, J. T., Herbel-Eisenmann, B. A., & Lloyd, G. M. (Eds.), *Mathematics teachers at work. Connecting curriculum materials and classroom instruction* (pp. 17–36). New York: Routledge.
- Brown, M. W., & Edelson, D. C. (2003). *Teaching as design: Can we better understand the ways in which teachers use materials so we can better design materials to support their changes in practice?* LeTUS Report Series : Evanston, IL.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Buczynski, S., & Hansen, B. (2010). Impact of professional development on teacher practice: Uncovering connections. *Teaching and Teacher Education*, 26(3), 599–607.
- Bybee, R. W. (2013). *The case for STEM education: Challenges and opportunities*. Arlington, VA: National Science Teachers Association.
- Capps, D. K., Crawford, B. A., & Conostas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *Journal of Science Teacher Education*, 23(3), 291–318.
- 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.
- Darling-Hammond, L., Hyster, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute. Retrieved January 1, 2020, from <https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report>
- Davis, E. A., Beyer, C., Forbes, C. T., & Stevens, S. (2011). Understanding pedagogical design capacity through teachers' narratives. *Teaching and Teacher Education*, 27, 797–810.
- Davis, E. A., & Varma, K. (2008). Supporting teachers in productive adaptation. In Kali, Y., Linn, M. C., & Roseman, J. E. (Eds.), *Designing coherent science education* (Vol. 5, pp. 94–122). New York: Teachers College Press.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Desimone, L. M. (2011). A primer on effective professional development. *Phi Delta Kappan*, 92(6), 68–71.
- Doppelt, Y., D., S. C., Silk, E. M., Mehlik, M. M., Reynolds, B., & Ward, E. (2009). Evaluating the impact of facilitated learning community approach to professional development on teacher practice and student achievement. *Research in Science and Technological Education*, 27(3), 339–354.
- Dubinsky, J. M., Roehrig, G. H., & Varma, S. (2013). Infusing neuroscience into teacher professional development. *Educational Researcher*, 42(6), 317–329.
- Dumais, N., & Hasni, A. (2009). High school intervention for influenza biology and epidemics/pandemics: Impact on conceptual understanding among adolescents. *CBE—Life Sciences Education*, 8, 62–71.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.

- Enderle, P., Dentzau, M., Roseler, K., Southerland, S., Granger, E., Hughes, R., ... & Saka, Y. (2014). Examining the influence of RETs on science teacher beliefs and practice. *Science Education*, 98(6), 1077–1108.
- Fischer, C., Eisenkraft, A., Fishman, B., Hubner, N., & Lawrenz, F. (2018). Adapting to the large-scale advanced placement chemistry reform: An examination of teachers' challenges and instructional practices. *Journal of Chemical Education*, 95(10), 1701–1710.
- Forbes, C. T., & Davis, E. A. (2010). Beginning elementary teachers' curriculum design and development of pedagogical design capacity for science teaching: A longitudinal study. In Kattington, L. E. (Ed.), *Handbook of curriculum development. Education in a competitive and globalizing world* (pp. 209–232). New York: Nova Science Publishers.
- Friedrichsen, P. J., & Barnett, E. (2018). Negotiating the meaning of *Next Generation Science Standards* in a secondary biology teacher professional learning community. *Journal of Research in Science Teaching*, 55(7), 999–1025.
- Gage, G. J. (2019). The case for neuroscience research in the classroom. *Neuron*, 102(5), 914–917.
- Greenleaf, C. L., Hanson, T. L., Rosen, R., Boscardin, D. K., Herman, J., & Schneider, A. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal*, 48(3), 647–717.
- Gross, P., Buttrey, D., Goodenough, U., Koertge, N., Lerner, L. S., Schwartz, M., & Schwartz, R. (2013). *Final evaluation of the Next Generation Science Standards*. Retrieved January 1, 2020, from <https://files.eric.ed.gov/fulltext/ED598819.pdf>
- Gustafsson, J.-E., Nilsen, T., & Yang, H. K. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, 57, 16–30.
- Hanuscin, D., Lipsitz, K., Cisterna-Alburquerque, D., Arnone, K. A., van Garderen, D., deArzuajo, Z., & Lee, E. J. (2016). Developing coherent conceptual storylines: Two elementary challenges. *Journal of Science Teacher Education*, 27, 393–414.
- Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3), 333–362.
- Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, 48(9), 590–598.
- Johnson, C. C., & Fargo, J. D. (2010). Urban school reform enabled by transformative professional development: Impact on teacher change and student learning of science. *Urban Education*, 45(1), 4–29.
- Johnson, C. C., & Fargo, J. D. (2014). A study of the impact of transformative professional development on Hispanic student performance on state mandated assessments of science in elementary school. *Journal of Elementary Science Teacher Education*, 25, 845–859.
- Kleickmann, T., Trobst, S., Jonen, A., Vehmeyer, J., & Moller, K. (2016). The effects of expert scaffolding in elementary science professional development on teachers' beliefs and motivations, instructional practices, and student achievement. *Journal of Educational Psychology*, 108(1), 21–42.
- Knight-Bardsley, A., & McNeill, K. L. (2016). Teachers' pedagogical design capacity for scientific argumentation. *Science Education*, 100(4), 645–672.
- Koster, M., & Bouwer, R. (2018). Describing multifaceted writing interventions: From design principles for the focus and mode of instruction to student and teacher activities. *Journal of Writing Research*, 10(2), 189–225.
- Kovarik, D. N., Patterson, D. G., Cohen, C., Sanders, E. A., Peterson, K. A., Porter, S. G., & Chowning, J. T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE—Life Sciences Education*, 12(3), 441–459.
- Kowalski, S. M., Taylor, J. A., Askinas, K. M., Wang, Q., Zhang, Q., Maddix, W. P., & Tipton, E. (2020). Examining factors contributing to variation in effect size estimates of teacher outcomes from studies of science teacher professional development. *Journal of Research on Educational Effectiveness*, 13(3), 430–458.
- Lara-Alecio, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The effect of an instructional intervention on middle school English learners' science and English reading achievement. *Journal of Research in Science Teaching*, 49(8), 987–1011.
- Lee, O., Deaktor, R. A., Hart, J. E., Cuevas, P., & Enders, C. (2005). Urban elementary schoolteachers' knowledge and practices in teaching science to English language learners. *Science Education*, 92(4), 733–758.
- Lontok, K. S., Zhang, H., & Daugherty, M. J. (2015). Assessing the genetics content in the Next Generation Science Standards. *PLoS ONE*, 10(7), 1–16.
- Lotter, C., Harwood, W. S., & Bonner, J. J. (2007). The influence of core teaching conceptions on teachers' use of inquiry teaching practices. *Journal of Research in Science Teaching*, 44(9), 1318–1347.
- Lotter, C., Smiley, W., Thompson, S., & Dickenson, T. (2016). The impact of a professional development model on middle school science teachers' efficacy and implementation of inquiry. *International Journal of Science Education*, 38(18), 2712–2741.
- Lynch, S., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260–293.
- MacNabb, C., Schmitt, L., Michlin, M., Harris, I., Thomas, L., Chittendon, D., ... & Dubinsky, J. M. (2006). Neuroscience in middle schools: A professional development and resource program that models inquiry-based strategies and engages teachers in classroom implementation. *CBE Life Sciences Education*, 5(2), 144–157.
- Marek, E., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research in Science Teaching*, 28(1), 41–53.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. V. (2004). Inquiry-based science in the middle grades: Assessment of learning in the urban systemic reform. *Journal of Research in Science Teaching*, 41(10), 1063–1080.
- McNeill, K. L., & Krajcik, J. S. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, 45(1), 53–78.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Nagle, B. (2013). Preparing high school students for the interdisciplinary nature of modern biology. *CBE—Life Sciences Education*, 12, 144–147.
- National Research Council (NRC). (2002). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education*. Washington DC.
- NRC. (2012). A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. *Committee to Develop a Conceptual Framework for New Science Education Standards Board on Science Education, National Research Council of the National Academies*. The National Academy Press: Washington, DC.
- Next Generation Science Standards Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press. Retrieved January 1, 2020, from www.nextgenscience.org/
- Next Generation Science Standards Lead States. (2013). *Next Generation Science Standards*. Washington, DC: National Academies Press.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84.
- Papaevripiidou, M., Irakleous, M., & Zacharia, Z. C. (2017). Using teachers' inquiry-oriented curriculum materials as a means to examine their pedagogical design capacity and pedagogical content knowledge for inquiry-based learning. *Science Education International*, 28(4), 271–292.
- 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.
- Perez, D., Van der Stuyft, P., del Carmen Zabala, M., Castro, M., & Lefevre, P. (2016). A modified theoretical framework to assess implementation fidelity of adaptive public health interventions. *Implementation Science*, 11, 91.
- Powell, J. C., & Anderson, R. D. (2002). Changing teachers' practice: Curriculum materials and science education reform in the USA. *Studies in Science Education*, 37, 107–113.

- Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13(2), 85–116.
- Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211–246.
- Remillard, J. T. (2018). Mapping the relationship between written and enacted curriculum: Examining teachers' decision making. In Kaiser, G., Forgasz, H., Graven, M., Kuzniak, A., Simmt, E., & Xu, B. (Eds.), *Invited lectures from the 13th International Congress on Mathematical Education*. (pp. 483–500). <https://doi.org/10.1007/978-3-319-72170-5>
- Roblin, N. P., Schunn, C., & McKenney, S. (2017). What are critical features of science curriculum materials that impact student and teacher outcomes? *Science Education*, 102(2), 1–23.
- Roehrig, G. H., & Garrow, S. T. (2007). The impact of teacher classroom practices on student achievement during the implementation of a reform-based chemistry curriculum. *International Journal of Science Education*, 29(14), 1789–1811.
- Roehrig, G. H., Kruse, R. A., & Kern, A. L. (2007). Teacher and school characteristics and their influence on curriculum implementation. *Journal of Research in Science Teaching*, 44, 883–907.
- Roehrig, G. H., Michlin, M., Schmitt, L., MacNabb, C. J., & Dubinsky, J. M. (2012). Teaching neuroscience to science teachers: Facilitating the translation of inquiry-based teaching instruction to the classroom. *CBE—Life Sciences Education*, 11(4), 413–424.
- Romine, W. L., Miller, M. E., Knese, S. A., & Folk, W. A. (2016). Multilevel assessment of middle school students' interest in the health sciences: Development and validation of a new measurement tool. *CBE—Life Sciences Education*, 15, ar21.
- Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwillie, K., & Wickler, N. I. Z. (2011). Video based lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, 48(2), 117–148.
- Roth, K. J., Wilson, C. D., Taylor, J. A., Stuhlsatz, M. A. M., & Hvidsten, C. (2019). Comparing the effects of analysis-of-practice and content-based professional development on teacher and student outcomes in science. *American Educational Research Journal*, 56(4), 1217–1253.
- Salli, U., Long, S. W., Carlsen, W. S., & Vrana, K. E. (2007). Stem cell biology should be taught in high schools. *CBE—Life Sciences Education*, 6, 283–284.
- Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht, Netherlands: Kluwer Academic.
- Schneider, R. M., & Krajcik, J. S. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education*, 13(3), 221–245.
- Schwartz, M. S., Hinesley, V., Chang, Z., & Dubinsky, J. M. (2019). Neuroscience knowledge enriches pedagogical choices. *Teaching and Teacher Education*, 83, 87–98.
- Schwartz, M. S., Sadler, P. M., Sonnert, G., & Tai, R. H. (2009). Depth versus breadth: How content coverage in high school science courses relates to later success in college science coursework. *Science Education*, 93(5), 798–826.
- Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: Investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies*, 41(4), 467–500.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453.
- Skultety, L., Gonzalez, G., & Vargas, G. (2017). Using technology to support teachers' lesson modifications during lesson study. *Journal of Technology and Teacher Education*, 25(2), 185–213.
- Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, 51, 870–901.
- Stark, J. C., Huang, A., Hsu, K. J., Dubner, R. S., Forbrook, J., Marshalla, S., ... & Jewett, M. C. (2019). BioBits health: Classroom activities exploring engineering, biology, and human health with fluorescent readouts. *ACS Synthetic Biology*, 8(5), 1001–1009.
- Straus, K. M., & Chudler, E. H. (2016). Online teaching resources about medicinal plants and ethnobotany. *CBE—Life Sciences Education*, 15(4), fe9.
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37, 963–980.
- Tashakkori, A., Teddlie, C., Johnson, B., & Gray, R. (2015). A history of philosophical and theoretical issues for mixed methods research. In Tashakkori, A., & Teddlie, C. (Eds.), *Sage handbook of mixed methods in social & behavioral research* (pp. 69–94). Thousand Oaks, CA: Sage.
- Taylor, J. A., Getty, S. R., Kowalski, S., 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.
- Taylor, J. A., Kowalski, S. M., Polanin, J. R., Askinas, K., Stuhlsatz, M. A., Wilson, C. D., & Wilson, S. J. (2018). Investigating science education effect sizes: Implications for power analyses and programmatic decisions. *AERA Open*, 4(3). <https://doi.org/10.1177/2332858418791991>
- Taylor, J. A., Roth, K., Wilson, C., Stuhlsatz, M., & Tipton, E. (2017). The effect of an analysis-of-practice, video case-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness*, 10(2), 241–271.
- Taylor, M. W. (2012). Replacing the "teacher-proof" curriculum with the "curriculum-proof" teacher: Toward more effective interactions with mathematics textbooks. *Journal of Curriculum Studies*, 45(3), 295–321.
- Van Dusen, B., & Nissen, J. (2019). Modernizing use of regression models in physics education research: A review of hierarchical linear modeling. *Physical Review Physics Education Research*, 15(020108). <https://doi.org/10.1103/PhysRevPhysEducRes.15.020108>
- Vazquez, J. (2006). High school biology today: What the committee of ten did not anticipate. *CBE—Life Sciences Education*, 5(1), 29–33.
- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340, 310–313.
- Windschittl, M. A., & Stroupe, D. (2017). The three-story challenge: Implications of the *Next Generation Science Standards* for teacher preparation. *Journal of Teacher Education*, 68(3), 251–261.
- Wood, W. B. (2002). Advanced high school biology in an era of rapid change: A summary of the biology panel report from the NRC committee on Programs for Advance Study of Mathematics and Science in American High Schools. *Cell Biology Education*, 1, 123–127.
- Yang, X., Hartman, M. R., Hattington, K. T., Etson, C. M., Fierman, M. B., Slonim, D. K., & Walt, D. R. (2017). Using next-generation sequencing to explore genetics and race in the high school classroom. *CBE—Life Sciences Education*, 16, ar22. [10.1187/cbe.1116-1109-0281](https://doi.org/10.1187/cbe.1116-1109-0281)
- Yang, Y., Liu, X., & Gardella, J. A. (2020). Effects of a professional development program on science teacher knowledge and practice, and student understanding of interdisciplinary science concepts. *Journal of Research in Science Teaching*, 57(7), 1028–1057.
- Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774), 364–369.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. US Department of Education: Washington, DC.
- Yoon, S. A., & Klopfer, E. (2006). Feedback (F) fueling adaptation (A) network growth (N) and self-organization (S): A complex systems design and evaluation approach to professional development. *Journal of Science Education and Technology*, 15(5), 353–366.
- Young, D. J., Reynolds, A. J., & Walberg, H. J. (1996). Science achievement and educational productivity: A hierarchical linear model. *Journal of Educational Research*, 89(5), 272–278. [10.1080/00220671.1996.9941328](https://doi.org/10.1080/00220671.1996.9941328)