Received: 29 February 2016 | Revised: 17 June 2017 | Accepted: 20 June 2017
DOI: 10.1002/sce.21302

WILEY

LEARNING

How much professional development is needed with educative curriculum materials? It depends upon the intended student learning outcomes

Anita M. Schuchardt[1] | Miray Tekkumru-Kisa[2] |
Christian D. Schunn[3] | Mary Kay Stein[3] | Birdy Reynolds[3]

[1]Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, MN, USA

[2]College of Education and Learning Systems Institute, Florida State University, Tallahassee, Fl, USA

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

Correspondence
Anita M. Schuchardt, Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, MN 55455, USA.
Email: aschucha@umn.edu

Abstract

There is little consensus on the kinds and amounts of teacher support needed to achieve desired student learning outcomes when mathematics is inserted into science classrooms. When supported by educative curriculum materials (ECM) and heavy investment in professional development (PD), teachers implementing a unit designed around mathematical modeling of scientific mechanisms substantially increased students' ability to make both qualitative and quantitative predictions (Schuchardt & Schunn, 2016). Because of concerns about equitable access to support resources, we investigated whether variations in PD support while retaining ECM could differentially affect two student learning outcomes: Quantitative Predictions and Qualitative Predictions. Two contrasts were performed examining: (1) the effect of reducing PD and (2) whether eliminating PD entirely caused further harm to student learning. Reducing and eliminating PD had no significant effect on student gains in Qualitative Predictions, suggesting ECM can be sufficient for teachers to support student learning of conceptual science content. However, student gains in Quantitative Predictions decreased significantly upon reducing PD; eliminating PD did not cause significant additional decreases. Combined, these findings suggest that amount of face-to face PD support necessary to achieve student-learning gains can vary depending on whether the practice requires application of qualitative science content or quantitative reasoning.

KEYWORDS
educative curriculum materials, mathematical modeling, Next Generation Science Standards, professional development, student learning

1 | INTRODUCTION

The Next Generation Science Standards (NGSS) places a prominent emphasis on mathematics: "Mathematics is a tool that is key to understanding science" (NGSS Lead States, 2013). Throughout the scientific practices identified in the NGSS (e.g., designing solutions, making predictions, or developing explanations), there is at least an implicit requirement for integrating mathematics with the science content. Unfortunately, while integrating mathematics with science education has been under investigation for the last century (Berlin & Lee, 2005), there is still little consensus on what is meant by integration and the effect of that integration on student learning. Two meta-analyses found a small average benefit for mathematics and science learning of integrating mathematics and science education (Becker & Park, 2011; Hurley, 2001). However, individual studies within the meta-analyses showed negative, positive, and neutral effects. The studies varied on how students were assessed, how mathematics was included in science instruction, and on the teacher supports provided. Because of the variation in the research literature on integrating mathematics into science education, there is little consensus on what types of mathematics integration facilitates student learning of science and the kind and amount of teacher support needed to produce that learning.

Recently, Schuchardt and Schunn (2016) found that, compared to traditional instruction, instruction that integrated mathematics by asking students to construct mathematical models of a biological phenomenon resulted in gains in students' conceptual understanding of the phenomenon as well as their ability to make quantitative predictions related to the phenomenon. That is, there were gains in students' learning in science tasks requiring mathematics as well as science tasks not needing mathematics. Those improvements occurred with teachers who had educative curriculum materials (Ball & Cohen, 1996) and were supported with a substantial amount of professional development (PD). In this paper, we investigate whether student benefits from this mathematics modeling approach still occur when teachers receive the same educative materials but face-to-face PD is reduced to more scalable levels.

1.1 | Inserting mathematics in science instruction

Historically, mathematics has been incorporated into science curriculum in four different ways: as an inscription, as a tool, as a grounded representation, or as a model (Schuchardt, 2016). Mathematics-as-inscription is often used when the goal is to convey how the scientific community does science. Instruction in graphing where students are lectured on the elements of a graph and commonly accepted graphing procedures is a prime example of this approach (Roth, Tobin, & Shaw, 1997). Mathematics-as-tool encompasses procedural approaches to mathematics where the goal of instruction is to solve an equation to achieve an answer. Students are taught a procedure to follow. Examples of this type of inclusion of mathematics include the probability rules equations in inheritance (Figure 1B), percent yield calculations in chemistry, and calculation of displacement given acceleration and initial velocity in physics. The common feature of mathematics-as-tool and mathematics-as-inscription is that connections to the underlying scientific phenomenon are masked or ignored. When this happens, student learning often fails to generalize beyond the specific use in initial instruction and is often misapplied in unfamiliar situations that require slight changes to the procedure (Stewart, 1983; Taasoobshirazi & Glynn, 2009; Tuminaro & Redish, 2007).

A third approach grounds the mathematical representations to aspects of the phenomenon. When students make those connections spontaneously (e.g., associate the cancelling out of two negatives in an equation with the experience of increased pressure with greater depth, or recognize that an equation is not just solving for some variable, but the amount of heat that water gains from a hot object), they are able to be more effective and flexible problem solvers (Bing & Redish, 2008; Gupta & Elby, 2011; Taasoobshirazi & Glynn, 2009; Tang, Tan, & Yeo, 2011). A few curricula in recent years try to create this grounding by having students construct their own graphical and mathematical representations of phenomenon. These curricular efforts are consistent with the most recent policies in science education that emphasize engaging students in disciplinary practices as they try to make sense of phenomena (Krajcik, Codere, Dahsah, Bayer, & Mun, 2014; National Research Council, 2012). Across different curricula, students have been shown to have greater understanding of the mathematical and graphical representations (Lehrer & Schauble, 2004; Roth &

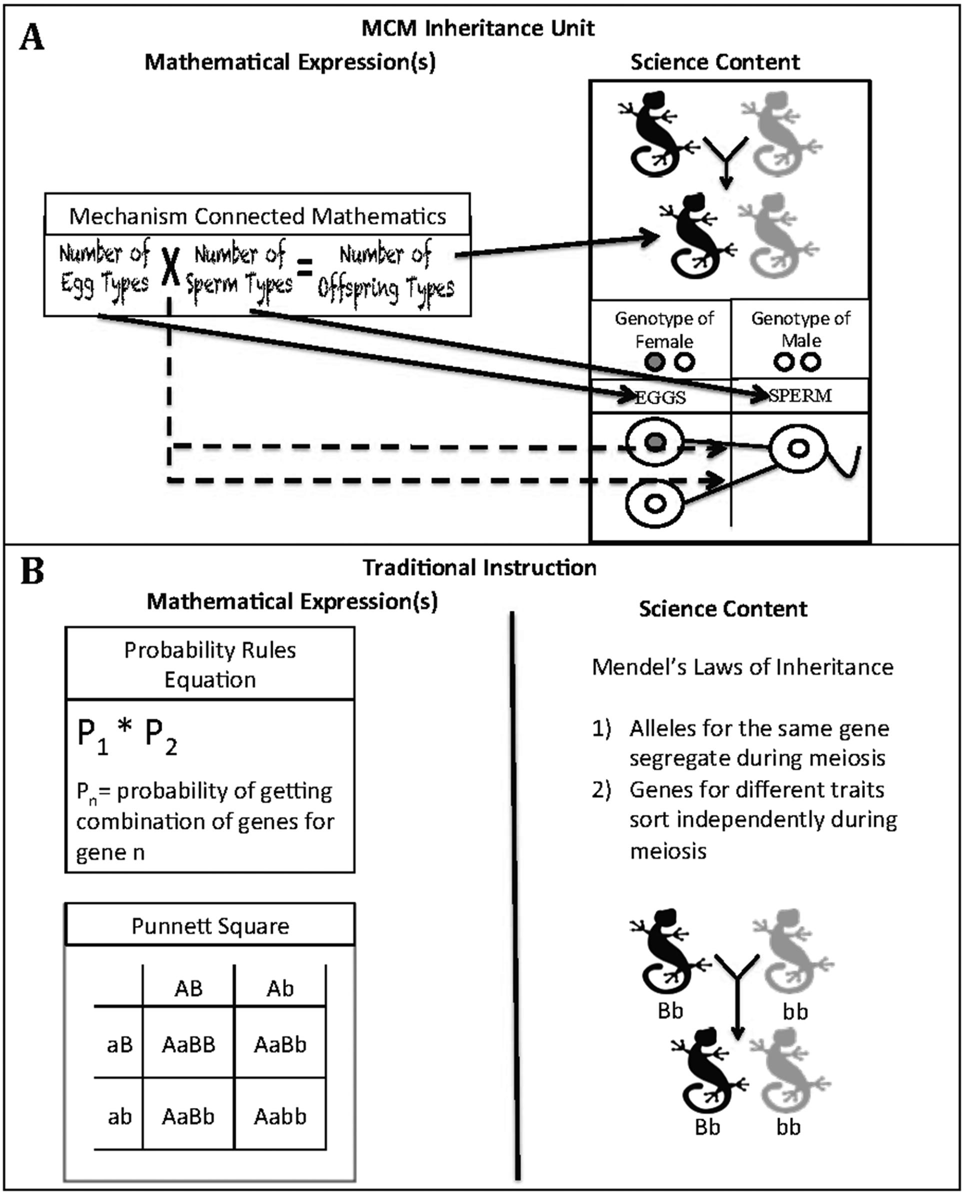


FIGURE 1 Relationship between mathematical expressions and science content in the MCM inheritance unit (A) and traditional instruction (B). In (A), the solid lines represent connections between entities in the phenomenon and variables in the mathematical expression, the dotted lines represent connections between the mathematical function of multiplication and the scientific mechanism of fertilization. In (B), the solid vertical line represents the common pattern of teaching mathematical expressions and science content in separate silos

Bowen, 1994; Wu & Krajcik, 2006), greater understanding of the represented scientific concepts (Levy & Wilensky, 2009; Malone, 2008; Wells, Hestenes, & Swackhamer, 1995), and greater flexibility in problem solving (Malone, 2008).

However, mathematical representations that students create or are provided with may not accurately depict the scientific mechanisms (Bechtel & Abrahamsen, 2005) that connect entities in the phenomenon. Equation development and use can become simple curve fitting exercises that generate equations convenient for calculation, not the equation forms that more closely correspond to the underlying scientific mechanisms (Schuchardt, 2016). There has been little exploration of the effect on student learning of mathematical representations that more directly include the scientific mechanism.

In a recent study, Schuchardt and Schunn (2016) investigated a high school biology curriculum unit (which we will call the Mechanism Connected Mathematics (MCM) unit) in which students construct a mathematical model of the scientific phenomenon of inheritance to determine the probability of getting a particular offspring when two parents breed (see Figure 1). Students were specifically asked to represent not just the entities of the phenomenon (sperm and egg) in their mathematical model, but also the mechanism (fertilization) (Figure 1A). It is this inclusion of the mechanism as well as the entities of the phenomenon that distinguishes a model from a grounded representation. It was theorized that by asking students to construct a mechanism-connected mathematical model from data, students would better understand the connections between their mathematical equation and the underlying mechanisms and entities of the biological phenomenon. This improved understanding would allow them to be better able to solve complex and novel quantitative problems and enhance their ability to apply their qualitative understanding of the biological process to make qualitative predictions. Schuchardt and Schunn (2016) found that, compared to traditional instruction that used an algorithmic approach to mathematics (Figure 1B), this type of inclusion of mathematics in science instruction resulted in a five-fold greater gain in students' understanding of the mathematically-modeled inheritance processes, and a four-fold greater gain in students' ability to solve complex inheritance problems. In other words, students were better able to make both qualitative predictions (e.g., what types of offspring) and quantitative predictions (e.g., the probability of each type of offspring) involving inheritance concepts. In this study, we are examining the supports that are necessary for teachers to successfully engage their students in constructing mechanism connected mathematical models of the phenomenon of inheritance.

The two scientific practices that improved (applying conceptual understanding of underlying scientific mechanisms to make predictions and solving quantitative prediction problems) have distinct functions that are fundamental to the pursuit of science. According to NGSS, "Students can be expected to evaluate and refine models through an iterative cycle of comparing their predictions with the real world and then adjusting them to gain insights into the phenomenon being modeled" (NGSS Lead States, 2013). Solving quantitative problems permits generation of a numerical prediction against which phenomenological outcomes can be rigorously tested, allowing for evaluation of whether a significant mismatch exists between the model and the phenomenon. Applying conceptual understanding of scientific mechanisms permits generation of qualitative predictions (what is observed) that can lead to identification of the nature of the mismatches between the conceptual model and the phenomenon that can lead to a more refined model of the phenomenon. In other words, the quantitative prediction permits evaluation of the significance of a mismatch between model and phenomenon, whereas the qualitative prediction permits identification of the nature of the mismatch. Because both of these practices are fundamental to science, in the current study, we continue to examine students' abilities with both practices.

1.2 | Supporting teachers' learning to include mathematics in science instruction

In early investigations of the MCM inheritance unit, it was found that high school biology teachers were underprepared to instruct students in mathematical modeling of the phenomenon (Cox, Reynolds, Schuchardt, & Schunn, 2016). Two general issues of pedagogical preparation were likely at play. One, as typically found in all areas of reform science instruction, teachers need PD to lead lessons that focus on student learning through engagement in the scientific practices (National Research Council, 2015). Two, more specific to this kind of instruction, teachers are often specifically underprepared to teach mathematics (Furner & Kumar, 2007; Offer & Mireles, 2009; Sorgo, 2010). Many science teachers have little background in mathematics (National Research Council, 2015) and they have usually only been exposed to science instruction where mathematics was included as algorithmic procedures to be memorized (Dancy & Henderson, 2010; Watanabe & Huntley, 1998). Therefore, when asked to think about ways to include mathematics in science instruction, they tend to generate mathematics as inscription or mathematics as tool approaches (Lee, Chauvot, Vowell, Culpepper, & Plankis, 2013; Offer & Mireles, 2009).

To address teachers' concerns by fully preparing them for implementing the MCM inheritance unit, the authors' initial study (2016) provided teachers with twenty-three hours of face-to-face PD on curriculum content and pedagogical practices (i.e., approximately five full days). Professional development generally requires an investment of physical, financial, and human resources in the form of time, money, space, and workshop leaders. Unfortunately, many educational systems are not equipped to provide large amounts of those resources (Spillane, Gomez, & Mesler, 2009). In our decades of experience with districts around the United States, school districts typically only provide one full day of PD in a given year to support teachers in a given topic. Most common is no PD at all. Twenty-three hours represent a heavy investment for even a well-resourced district. In moving this kind of NGSS-aligned curriculum to multiple contexts, and thus, providing equitable access, efficient amounts of PD need to be found that elevate student learning while minimizing the investment of resources by local agencies.

1.3 | Professional development with educative curriculum materials

One approach that has been suggested for taking innovative curricula to scale while minimizing investment of local resources is to provide educative curriculum materials along with face-to-face PD. Davis and Krajcik (2005) suggest teacher educative curriculum materials can help teachers learn content, likely student responses to instructional activities, the relationships between units, the designers' rationale behind activities, and understanding of new pedagogies. Research suggests that teachers who use educative curriculum materials do show changes in their instruction, including using a greater number and more varied strategies to support learning and changes to teacher Pedagogical Content Knowledge (PCK; Cervetti, Kulikowich, & Bravo, 2015; Schneider, 2013). Very few studies in science education have examined the effect on student learning of giving teachers educative curriculum materials. These few exceptions have been situated in upper elementary school (grades 3rd through 6th, ages 8 through 11) (Arias, Smith, Davis, Marino, & Palincsar, 2017; Cervetti et al., 2015; Lin, Lieu, Chen, Huang, & Chang, 2012). Two studies have explored the effect of the presence or absence of educative curriculum materials either when all teachers in the study have face-to-face PD (Arias et al., 2017) or when all teachers are not provided with face-to-face PD (Cervetti et al., 2015).

Although educative curriculum materials are likely more effective in conjunction with other forms of support, such as PD (Cervetti et al., 2015; Davis & Krajcik, 2005), these materials will often be used without this kind of support. Therefore, in the current study, we look at the effect on the learning of older students when teachers are provided with different amounts of face-to-face PD, all in the context of educative curriculum materials. Knowing the effect of variations in teacher supports will help policy makers, curriculum designers, and administrators make decisions about investments in these additional supports as NGSS curricula moves to scale.

1.4 | The importance of investigating the effect of PD variations on student outcomes

This current investigation of the role of amount of teacher support focuses on student learning outcomes rather than improvements in teaching. Although effects on teaching are clearly important and more commonly investigated in studies of PD (Dede, Ketelhut, Whitehouse, Breit, & McCloskey, 2009; Doppelt et al., 2009; Luft & Hewson, 2014; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), it is not necessarily the case that changes in particular teacher beliefs or practices necessarily produce improvements in student outcomes (Kleickmann, Trobst, Jonen, Vehmeyer, & Moller, 2016) and it is therefore useful to examine effects on students. In science, in particular, very few studies on PD have looked at student learning gains (Scher & O'Reilly, 2009; Wilson, 2013; Yoon et al., 2007) and some of these studies confound PD effects with intervention effects (e.g., Marek & Methven, 1991; Radford, 1998). Moreover, the effect of varying teacher support on student outcomes for these two different predictive practices is under-investigated. Often, researchers measure general science content knowledge (e.g., Diamond, Maerten-Rivera, Rohrer, & Lee, 2014; Doppelt et al., 2009; Radford, 1998) or overall knowledge about specific units (e.g., Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Taylor, Roth, Wilson, Stuhlsatz, & Tipton, 2016).

In sum, we take on the under-studied research question of whether variations in face-to-face PD support for teachers implementing a mathematical modeling unit can differentially affect two student learning outcomes associated

TABLE 1 Comparing Traditional and MCM Instruction

	Instructional form	
Instructional characteristic	**Traditional instruction**	**MCM instruction**
Mathematical equations	Mathematical equations and procedures provided	Mathematical equations generated by students
Integration of mathematics and science	Mathematics and science taught in parallel silos	Mathematics and science processes taught simultaneously
Expectation of students	Students study content provided by teacher and textbook	Students develop content understanding through analyzing data, developing and refining models, arguing from evidence

with scientific practice: Quantitative Predictions and Qualitative Predictions. More specifically, across three conditions of face-to-face teacher PD in the context of the same instructional materials and the same educative curriculum materials (i.e., the MCM inheritance unit), what is the effect on student learning outcomes in quantitative predictions and qualitative predictions? We contrast: (1) the full amount of PD, (2) a greatly reduced amount that district officials typically assign to address PD needs (i.e., one full day), and 3) the amount of teacher PD that typically happens in the United States without special allocation of resources (i.e., none).

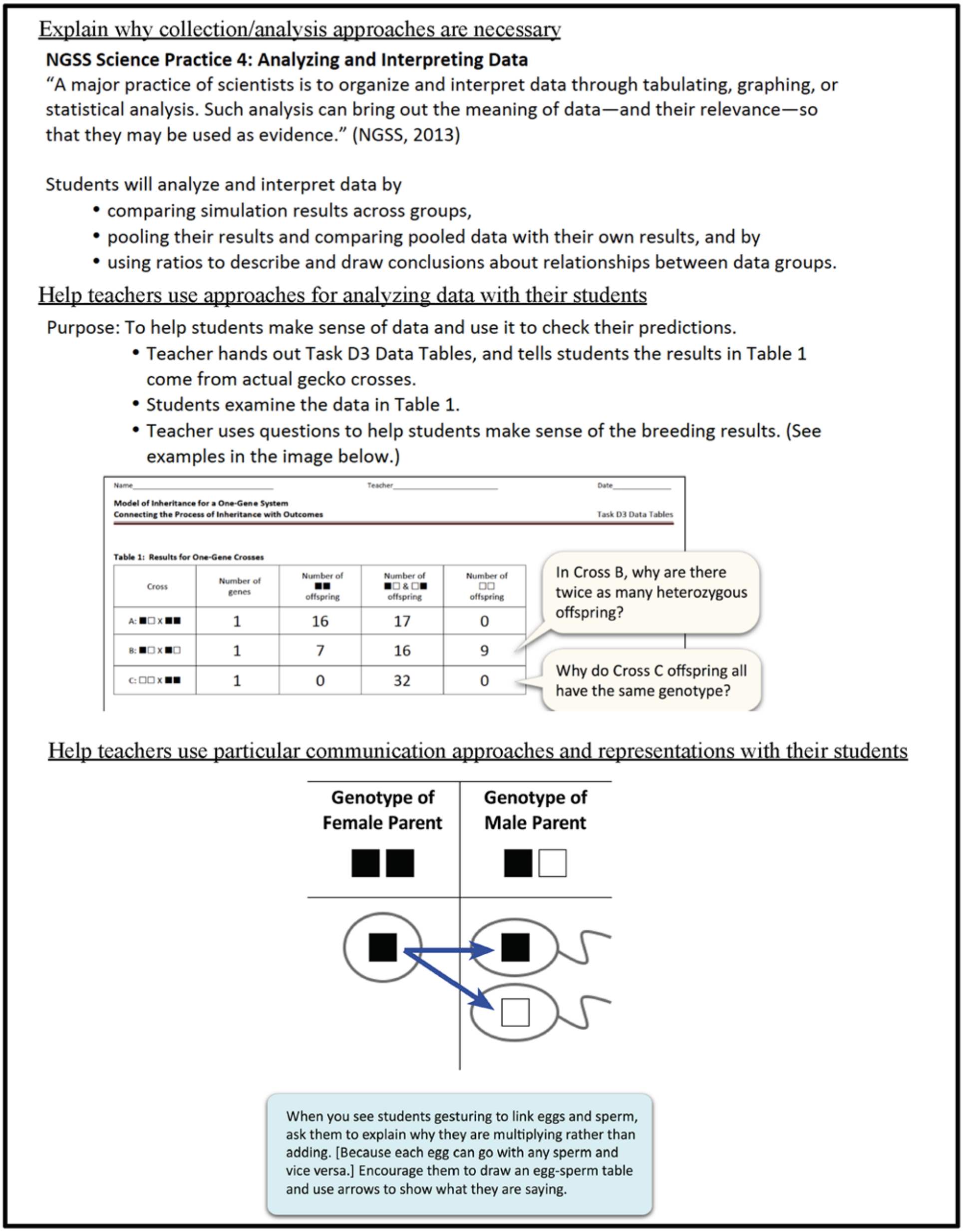
2 | METHODS

The first section of the methods discusses three elements that were the same across all three variations in face-to-face PD: the MCM inheritance unit (which includes the educative curriculum materials and the student instructional materials) and the student assessment.

2.1 | MCM inheritance unit

The description of the MCM unit contained here focuses on those details that illustrate how the MCM inheritance unit differed from traditional instruction in inheritance (see Schuchardt & Schunn, 2016 for details). This unit was designed to incorporate a type of insertion of mathematics into science that treated mathematics as a model. The MCM inheritance unit asked students to develop mathematical models (defined as those mathematical representations that include both the entities of the phenomenon as well as the scientific mechanism that connects the entities) of genetic processes in an iterative cycle. As written, the MCM inheritance unit contained both differences in practices and differences in scientific content compared to how genetics has traditionally been taught (see Table 1).

The science content covered in the unit was also subtly but importantly different from the content contained in traditional curriculum, given the NGSS-based emphases on greater depth, connections to mathematics, and development of content through engagement in scientific practices (Table 1 and Figure 1). First, instead of learning inheritance laws as rote facts, the MCM unit asked students to develop inheritance laws as they make sense of and mathematically model data on inheritance outcomes. Second, students were asked to generate mathematical models of the phenomenon by making explicit connections between their mathematical equations and their understanding of inheritance entities and processes. Third, the analysis and modeling activities were designed so that students were asked to develop an understanding of probability in inheritance as the proportion of the total outcome space occupied by the event space (e.g., how many of the desired offspring types (event space) as a proportion of all possible offspring from the given parents (outcome space)). This contrasted with the focus in traditional inheritance instruction on memorizing the laws for combining probabilistic events (e.g., “AND” always implies multiply). The proportional representation of probability served to explicitly tie the mathematical idea of probability to categories (desired offspring, all possible offspring) within the phenomenon.



Explain why collection/analysis approaches are necessary

NGSS Science Practice 4: Analyzing and Interpreting Data

"A major practice of scientists is to organize and interpret data through tabulating, graphing, or statistical analysis. Such analysis can bring out the meaning of data—and their relevance—so that they may be used as evidence." (NGSS, 2013)

Students will analyze and interpret data by

- comparing simulation results across groups,
- pooling their results and comparing pooled data with their own results, and by
- using ratios to describe and draw conclusions about relationships between data groups.

Help teachers use approaches for analyzing data with their students

Purpose: To help students make sense of data and use it to check their predictions.

- Teacher hands out Task D3 Data Tables, and tells students the results in Table 1 come from actual gecko crosses.
- Students examine the data in Table 1.
- Teacher uses questions to help students make sense of the breeding results. (See examples in the image below.)

Name________ Teacher________ Date________

Model of Inheritance for a One-Gene System
Connecting the Process of Inheritance with Outcomes

Task D3 Data Tables

Table 1: Results for One-Gene Crosses

Cross	Number of genes	Number of ■■ offspring	Number of ■□ & □■ offspring	Number of □□ offspring
A: ■□ x ■■	1	16	17	0
B: ■□ x ■□	1	7	16	9
C: □□ x ■■	1	0	32	0

In Cross B, why are there twice as many heterozygous offspring?

Why do Cross C offspring all have the same genotype?

Help teachers use particular communication approaches and representations with their students

Genotype of Female Parent	Genotype of Male Parent
■■	■□

When you see students gesturing to link eggs and sperm, ask them to explain why they are multiplying rather than adding. [Because each egg can go with any sperm and vice versa.] Encourage them to draw an egg-sperm table and use arrows to show what they are saying.

FIGURE 2 Excerpts from the educative curriculum materials, illustrating some of the supports provided to teachers [Color figure can be viewed at wileyonlinelibrary.com]

2.2 | Curricular materials provided to teachers

All teachers were provided with the same set of student materials (e.g., worksheets and manipulatives). All teachers also received access to the same set of educative curriculum materials, designed to provide information about student ideas and teacher pedagogical practices. (An example of the materials is shown in Figure 2.) To support this claim about the educative nature of the materials, a representative set of materials from six of the fourteen tasks (representing a mixture of conceptual development and mathematical modeling) were analyzed for educative properties using the criteria for educative quality from Beyer, Delgado, Davis, and Krajcik (2009) in their analysis of teacher supports for high school biology curriculum. First, each task was analyzed holistically to determine which teacher knowledge domain and category could be represented (as defined by Beyer et al.; see below for definitions and examples of each). Then the teacher materials were analyzed for each task to determine if the educative criteria for those knowledge domains were present (e.g., helping teachers use approaches for collecting and analyzing data, helping teachers use representations of scientific phenomenon with students). To increase independence and validity of results, the first author, who was familiar with the unit, but had not developed the teacher support materials, completed the coding.

The curricular materials addressed all three teacher knowledge domains from the Beyer et al. criteria: PCK for science topics, PCK for scientific inquiry, and teacher's subject matter knowledge. Further, they provided support in eight out of the nine specific categories, including: (a) engaging students with topic-specific scientific phenomena, (b) using scientific instructional representation, (c) anticipating and dealing with students' ideas about science, (d) engaging students in questions, (e) engaging students with collecting and analyzing data, (f) helping students make explanations based on evidence, (g) promoting scientific communication, and (h) development of subject matter knowledge (Beyer et al., 2009). The ninth category, supporting teachers in engaging students in designing science investigations, was not relevant as students did not design science investigations in this unit; instead they analyzed provided data and engineered solutions.

The educative curriculum materials were in print but also located online to provide a forum where teachers could ask questions, post student work, and share ideas. Teachers in all conditions were provided with equal time and training on how to access the online materials. This training was completed after PD on the curriculum was finished and teachers did not access the online educative curriculum materials as part of PD.

Our analysis showed that teachers logged in regularly to access the provided materials (a mean of 38 times, ranging from 16 to 80 times across teachers). There was little variation across PD conditions. Teachers in both face-to-face and online only conditions had access to each other's comments, questions and uploads. However, our login data collection did not allow us to determine exactly what teachers were doing when they logged in. Only a few of the teachers engaged in online discussions or posted revisions and those who did, did so infrequently. Therefore, while teachers in all conditions had access to this online platform, discussions between teachers formed only a small part of the online curriculum material.

2.3 | Student assessment

The focus of this investigation is the effect of different levels of PD on student prediction practices for both: (a) qualitative predictions and (b) quantitative predictions. Qualitative prediction questions are those that assess students' ability to make qualitative predictions about outcomes (e.g., the types of gene combinations that an offspring will have, Figure 3) by applying conceptual understanding of scientific mechanisms (covering the processes involved with transmission of genes between parents and offspring (meiosis and fertilization)). Quantitative prediction questions ask whether students can determine the probability of a particular outcome in a genetic context (e.g., the numerical probability that an offspring will contain a specific set of genes; see Figure 3). Schuchardt and Schunn (2016) found that students showed gains from the MCM unit in making quantitative predictions for questions involving complex genetic probability (two or more genes), but not simple genetic probability (one gene). Since the intent here was to see whether these gains are maintained when the amount of teacher support is reduced, only genetic probability questions involving two or more genes were included in this analysis.

The assessments were the same as used in Schuchardt and Schunn (2016), for comparability. The posttests had a mean KR-20 (Kuder and Richardson Formula 20 test, (Kuder & Richardson, 1937)) of 0.72 (mean discrimination = 0.46; mean difficulty = 0.50). (See Schuchardt and Schunn (2016) for additional details regarding test item sources, and on reliability and validity of the tests.)

Teachers administered the assessments before and after implementation of the MCM inheritance unit following a matrix sampling protocol to allow for broad coverage of the conceptual content but not consume two full class periods for testing. In other words, there were multiple pretest versions and multiple posttest versions that teachers distributed randomly within each of their classes. From this method of testing, analyses focused on composite scores

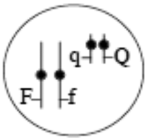
Question Category	Number of Questions	Example Question
Qualitative Predictions	7	The genotypes of the sperm from one male and the genotypes of the eggs from one female are shown below. Male Sperm Female Eggs FQ FQ fQ fQ Fq fq Which answer lists all the possible genotypes that could be expected in the offspring: a. FFQQ, ffQQ b. FFQQ, FfQQ, FFQq, FfQq, ffQQ, ffQq c. FFQQ, FfQQ, FfQq, ffQq d. FFQQ, FFQq, FFqq, ffQQ, ffQq, ffqq, FfQQ, FfQq, Ffqq
		In the germline cell below there are two pairs of chromosomes on which are shown the locations of two different genes. F and f represent two different alleles (versions or variants) of one gene, and Q and q represent two different alleles of another gene. If this cell divides <u>normally</u> to produce sperm, what are the possible sperm genotypes?  a. F, f, Q, q b. Ff, Ff, Qq, Qq c. FQ, fq, Fq, fQ d. Ff, Qq, FQ, fq, Fq, fQ
Quantitative Predictions	7	If organisms of type BbSs and type bbSs are crossed, what is the probability that the offspring would be BbSs? a. 1/16 b. 1/8 c. 1/4 d. 1/2 e. 9/16
		Given a female with the genes: AaBbCc, what proportion of her eggs will contain genes “a” AND “b”? a. 1/4 b. 3/8 c. 1/2 d. 2/3

FIGURE 3 Question categories on pre and post assessments. The number of questions refers to the number of questions in the pool

across students for each question rather than individual student scores aggregating across questions. Details of each specific statistical analysis procedure are presented within each relevant results section.

2.4 | Participants

We contrasted the performance of ninth- and tenth-grade teachers from Schuchardt and Schunn (2016) against performance of ninth- and tenth-grade teachers participating in the following year with one of two different levels of reduced PD. Students in these grades typically range in age from 14 to 16 years. Across the two cohorts, 24 teachers were recruited from primarily urban and suburban school districts surrounding two metropolitan areas located in midwestern states. All teachers were compensated for their participation in the study. Recruitment was done through a flyer distributed via regional instructional support organizations soliciting teachers to attend a 2-hour information session on implementing a unit in biology aligned with NGSS. During the first year of the study, after attending an information session that provided an overview of the unit, six teachers participated in all 23 hours of face-to-face PD sessions (henceforth called the 23Hour PD condition). During the following year, the 12 teachers who volunteered to

TABLE 2 Teacher, professional development content, and student characteristics of each professional development condition

Characteristic	23 Hour PD	8 Hour PD	No PD
Teachers			
Number of teachers	6	8	4
Number of teachers reporting educational information	6	7	4
Number of teachers with undergraduate or master's degree in biology	6	6	3
Number of teachers with master's degree in biology	1	1	0
Number of teachers with master's degree in education	1	4	2
Students			
Number of students	265	377	162
Mean percent of students qualifying for free and reduced lunch (SD)	32 (23)	55 (23)	61 (26)
Mean qualitative prediction pretest score (SD)	36 (4)	32 (7)	35 (5)
Mean quantitative prediction pretest score (SD)	38 (11)	32 (9)	29 (7)
Professional development			
Mean total number of hours of PD	23	8	0
Mean number of hours of PD on content	13	5	0
Mean number of hours of PD on pedagogy	10	3	0

implement participated in one of two conditions: no face-to-face PD (No PD, four teachers) or 8 hours of face-to-face PD (8Hour PD, eight teachers).

Although not random assignment to condition, conditions were designed to be balanced after taking into account district constraints. Three of the teachers in the 8Hour PD condition did not have the option to participate in the No PD condition because they were participating as part of a continuing education program for their regional educational organization. The remaining nine teachers were asked which condition they would prefer to be in when they signed up; next, changes were made in condition assignment (with teacher consent) to ensure a balance among these nine teachers by student and teacher characteristics across the 8Hour PD and No PD conditions. In all conditions, teachers were provided with the same educative curriculum materials.

Overall, all three groups were well matched based on teacher experience, teacher education, and school characteristics. Almost all of the teachers had either a master's or undergraduate degree in biology (see Table 2), and most had been teaching for 11 or more years. The student characteristics in Table 2 illustrate both the diversity of contexts studied and strongly overlapping distributions across conditions—in the United States, whether students qualify for free or reduced cost lunch is used as the primary indicator of socioeconomic status. Statistical analyses of the student data included student and school characteristics as control variables.

2.5 | Professional development conditions

In the 23Hour PD condition, all teachers received extended face-to-face PD consisting of a weeklong summer workshop and two follow-up sessions during implementation of the unit. In the 8Hour PD, teachers had only eight face-to-face PD hours across two sessions during the school year. (This analysis of time only included time spent on PD in the MCM inheritance unit. It did not include time spent on teacher surveys and assessments, administrative activities, or training in accessing the online materials.)

The face-to-face PD in both years of the study had the critical features of high-quality PD identified in the research literature: (a) active learning, (b) collective participation, (c) embedded in subject matter, and (d) coherent with current policy (e.g., NGSS) (Desimone, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001; Reiser, 2013; Wilson, 2013). The fifth characteristic, of sufficient duration, is what is being tested between years one and two: what counts as sufficient duration for which intended student learning outcomes in the context of educative curriculum materials.

During both years, teachers in the PD workshops engaged in the MCM inheritance unit as learners, participating in both small-group and whole-group discussions to develop their conceptual understanding of the material covered in the unit. These sessions were conducted in a way that was coherent with NGSS practices (National Research Council, 2012) with the workshop leaders acting as teachers and modeling the pedagogical practices that teachers would be expected to enact with their students. Two workshop leaders conducted the PD in the second year, and they were two of the four workshop leaders who conducted the PD in the first year. In both years, leaders had expertise in the biological sciences and pedagogy, participated in the design of the MCM inheritance unit, and had multiple years of prior experience leading extended PD workshops around reform instruction in science.

Based on our records of PD activities, the 23 hours of face-to-face PD during the first year of the study was divided between 10 hours primarily focused on pedagogy and 13 hours primarily focused on content (see Table 2). Following best practices in PD, all of the pedagogy-focused PD activities were embedded in the context of the unit. However, the pedagogy-focused PD activities were separated from the content-focused activities by the learning goals for the teachers. Pedagogy was defined as any activity where teachers were expected to learn about how to support student learning in the context of the MCM inheritance unit (i.e., the role of the engineering challenge, the relationship of the unit to NGSS and state standards, the role of multiple representations) or about teaching practices (i.e., facilitating student discourse, reflection on teaching, examining aligned and nonaligned enactments of unit instruction).

Teacher content-learning PD activities included 10 hours focused on having teachers learn about the content of the unit (especially related to both the conceptual explanations of biological inheritance processes and the quantitative prediction tasks), by experiencing the unit as learners. These content learning activities were approximately evenly divided between a focus on learning conceptual science content (the genetic processes of meiosis and fertilization) and quantitative problem solving (solving genetic probability problems). An additional 3 hours was spent on exposing teachers to common mistakes made by students when making quantitative predictions and helping teachers to reason through the types of conceptual errors that were likely to have caused these mistakes. While teachers were thinking about how students conceptualize inheritance in this PD task, they were also engaging heavily in assessment and reevaluation of their own understanding of content and thus, these 3 hours were grouped into the content learning activities.

The PD was shortened during the second year of the study by focusing on the most important elements as judged from teacher struggles and feedback during prior iterations: Developing content knowledge was reduced from 13 to 5 hours, and pedagogy was reduced from 10 hours to 3 hours (see Table 2). Both content and pedagogy training were carried out in the same way as in the prior year of the study. The 3 hours spent on determining the conceptual foundation of common errors in student quantitative problem solving were eliminated.

3 | RESULTS

3.1 | Equivalence of pretest scores across professional development conditions

The means and standard deviation of pretest scores by teacher for each student predictive practice (qualitative or quantitative) are shown in Table 2 and reflect the wide range of student backgrounds. Within each of the predictive practices, pretest scores were examined for statistically significant differences across the implementing conditions using a one way between-subjects analysis of covariance (ANCOVA) conducted on the pretest question means for each teacher's students, with PD condition as the independent variable and free and reduced lunch (FRL) as the covariate. All required statistical assumptions were met (e.g., normality, homogeneity of variance, outliers, and independence).

While a marker of socioeconomic status, percent of FRL students, did not show significantly different means across conditions, the standard deviations for each condition was quite large. Therefore, to err on the conservative side, FRL was kept as a covariate on the tests of differences in learning gains across conditions.

For qualitative predictions and quantitative predictions, pretest means adjusted for FRL were not significantly different across PD conditions ($F(2,14) = 0.11$, $p = .90$; $F(2,14) = .59$, $p = .57$). This analysis shows that students in the three groups (23Hour PD, 8Hour PD, and No PD) had similar content knowledge in inheritance prior to instruction.

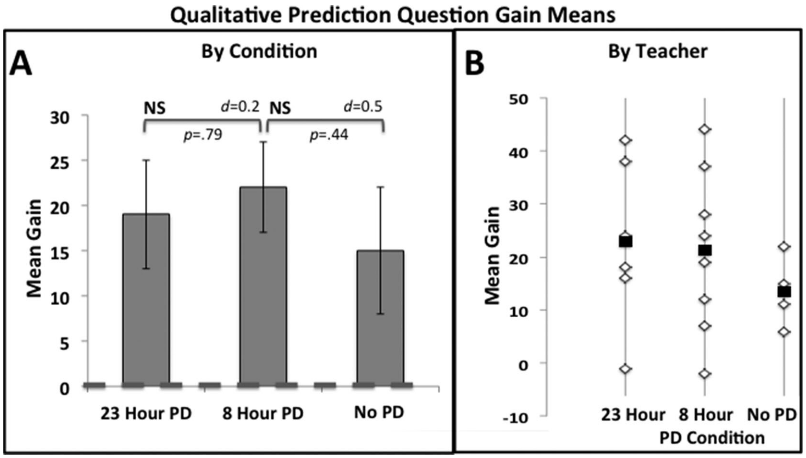


FIGURE 4 Pre–post gains in qualitative prediction. (A) Mean gains (with SE bars) by condition. Statistical significance and effect size details are presented for each planned contrast. The dotted line shows the mean pre–post change in each content area for a group of comparison teachers using a traditional curriculum (Schuchardt & Schunn, 2016). (B) Gains for each individual teacher's students by condition in open diamonds, and mean gain across teacher by condition in black squares

As noted earlier, because pre- and posttests involved matrix sampling, the data were analyzed for generalizability across questions rather than across students: A mean for each question was calculated for each teacher pre- and postinstruction. A gain score for each question was calculated by subtracting the preinstruction question mean (averaged across the students from each teacher) from the postinstruction question mean (averaged across the students from each teacher). These question gain scores were then aggregated within each predictive practice to provide a mean category gain score for each teacher. To remove the nuisance variance associated with initial differences in pretest scores across teachers, statistical analyses were conducted on these mean category gain scores.[1]

To examine consistency of gain scores across teachers within each PD condition, one-way between-subjects ANCOVAs (with FRL as covariate) were conducted on the teacher gain scores for each learning outcome. All assumptions were met (e.g., normality, homogeneity of variance, outliers, and independence). Two planned contrasts were performed for each learning outcome: (a) 8Hour PD against 23Hour PD to determine whether reducing the number of face-to-face PD hours for teachers affected student learning and (b) 8Hour PD against No PD to determine whether educative curriculum materials alone were sufficient to achieve student learning gains. In Figures 4A and 5A, the brackets indicate the contrasts that were performed. Statistical test results and effect sizes (Cohen's *d*) are presented above and below the brackets in the relevant figures.

3.2 | Effect of reducing face-to-face PD on qualitative predictions

Comparison of student learning gains on qualitative predictions in the 8Hour face-to-face PD condition to the 23Hour PD conditions revealed a small effect size that was not statistically significant ($F(1,14) = 0.07$, $p = .8$, 95% CI [−21, 16], Figure 4A, first bracket). Because students were able to achieve the same learning gains on qualitative predictions, this result suggests that students and teachers in the 8Hour face-to-face PD condition have the same capacity for learning gains as students and teachers in the 23Hour PD condition. In both PD conditions, student-learning gains were significant (mean gains of 19–22) and greater than the nonexistent gains that were achieved in the previous study with traditional curriculum (Schuchardt & Schunn, 2016). Thus, this result is useful for supporting the claims of equivalence of students and teachers across these two conditions.

To determine whether having any face-to-face PD support for teachers was necessary for student learning gains to occur in the MCM unit, student performance in the 8Hour PD group was compared to that in the No PD group (Figure 4A, second bracket). Decreasing face-to-face PD from 8 hours to 0 hours had a moderate effect on the mean

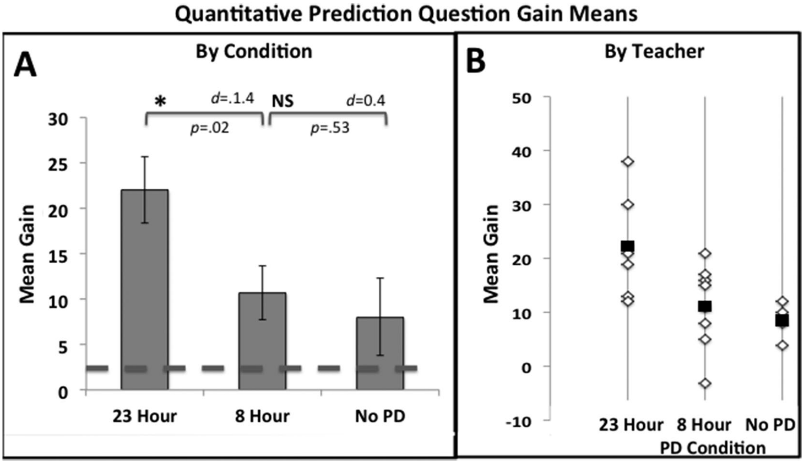


FIGURE 5 Pre–post gains in quantitative prediction. (A) Mean gains (with SE bars) by condition. Statistical significance and effect size details are presented for each planned contrast. The dotted line shows the mean pre–post change in each content area for a group of comparison teachers using a traditional curriculum (Schuchardt & Schunn, 2016). (B) Gains for each individual teacher's students by condition in open diamonds, and mean gain across teacher by condition in black squares

performance of the No PD group, but the mean was not significantly different from the 8Hour PD group ($F(1,14) = 0.6$, $p = .4$, 95% CI [–26, 11], Figure 4A) and the mean gain of 15 for the No PD group was still meaningful. Teacher means of student responses were evenly distributed around the condition mean in all three conditions (Figure 4B). Moreover, Schuchardt and Schunn (2016) had shown that students who received traditional instruction had shown no learning gains on this measure (dotted line in Figure 4A). Combined, these results suggest that educative curriculum materials alone were sufficient to produce most of the student learning gains from using an NGSS-aligned curriculum in qualitative predictions, but does not entirely rule out small additional beneficial effects of having some face-to-face PD.

3.3 | Effect of reducing face-to-face PD on quantitative predictions

In contrast to the lack of a significant effect on qualitative predictions of reducing PD from 23 hours to 8 hours, this reduction in PD support had a large effect on pre–post gains on quantitative predictions, resulting in a significant decrease in teacher means of student gains ($F(1,14) = 6.5$, $p = .02$, 95% CI [2, 23], Figure 5A, first bracket). For the 23Hour PD condition, student showed a nontrivial mean learning gain of 22, which had been shown in the prior study to be significantly greater that that achieved by traditional instruction. (For the sake of comparison, the mean gain for students who received traditional instruction is shown by the dotted line in Figure 5A; Schuchardt & Schunn, 2016). Teacher means of student responses in the 23Hour and 8Hour PD conditions were evenly distributed around the condition mean (Figure 5B), suggesting that the deleterious effect of reducing PD is not simply due to the effect of outliers. Comparison of the 8Hour PD group with the No PD group showed that removing all face-to-face instructional support had only a small effect and did not result in a significant further drop on learning of this predictive practice ($F(1,14) = 0.4$, $p = .5$, 95% CI [–14, 8], Figure 5A, second bracket). Again, teacher means of student responses were evenly distributed around the condition mean in both conditions (Figure 5B). These results suggest that in contrast to qualitative predictions, quantitative predictions by students are less robust to reductions in face-to-face PD.

4 | DISCUSSION

The results presented here suggest that when teachers are provided with educative curriculum materials to facilitate implementation of a unit that incorporates mathematical modeling, the amount of face-to-face PD necessary to support student-learning gains depends on the intended learning outcomes. For making qualitative predictions by applying core conceptual science content (i.e., the content most familiar to teachers), statistically equivalent student-learning gains can be achieved even in the absence of face-to-face PD (i.e., just with educative curriculum materials), although some face-to-face PD appears to be helpful. By contrast, for making quantitative predictions, which involves mathematical application in a science context and thus is more novel to teachers, student-learning gains are greatly reduced when face-to-face PD is reduced to 8 hours. The findings of this study, therefore, suggest that in the context of educative curriculum materials, when teachers have a lot of experience with the content (i.e., the biological mechanisms), then teachers need little if any face-to-face PD support. However, when teachers have less experience with the content (i.e., mathematical modeling), teachers need more face-to-face PD support, even in the context of educative curriculum materials. These results therefore have implications for implementation of NGSS with its emphasis on scientific practices, the design of teacher PD, and the importance of educative curriculum materials.

NGSS express the value of having students learn and engage in scientific practices as well as scientific content (NGSS Lead States, 2013). The current results suggest that students may achieve different levels of success with a single scientific practice learning outcome (making predictions) depending on the nature of the task (i.e., the ability to make quantitative as opposed to qualitative predictions). Furthermore, our results suggest that success with the two different contexts (quantitative vs. qualitative) of the same practice (making predictions) can be differentially affected by the amount of support provided to their teachers. Qualitative predictions that draw on knowledge about biological mechanisms are more robust to reductions in face-to-face PD than quantitative predictions that draw on an understanding of mathematical representations of biological mechanisms. Different student learning outcomes may require different amounts of support for teachers because teacher content knowledge is not monolithic (National Research Council, 2015). For example, in the context of this MCM inheritance unit, biology teachers are well-prepared to teach about the biological mechanisms of inheritance (Lyons, 2013). The unit requires no new understanding of this biology content, but rather involves a different way of helping their students learn this content. On the other hand, biology teachers are generally not as well versed in mathematics (Sorgo, 2010). Understanding and calculating probability in a new way in this science context would therefore likely require more support, and so would drawing connections between the functions of a mathematical representation and the biological mechanism. The different student learning outcomes effects observed here provide a different lens on the effect of teacher supports on student learning of scientific practices, each of which is discussed below.

As the implementation of a curriculum that incorporates NGSS practices such as mathematical modeling moves beyond field trials of curriculum materials to implementation at scale, it is encouraging that significant student learning gains were observed in the area of applying core science content (e.g., scientific mechanisms) to make qualitative predictions when teachers were supported by educative curriculum materials alone. These gains with minimal support are particularly impressive considering students in traditional instruction had previously shown no significant gains in their ability to make qualitative predictions based on these core scientific mechanisms that are at the heart of inheritance (i.e., this content is quite difficult for students; Schuchardt & Schunn, 2016). Note that the students in the no PD group were not specially prepared to master this material: (1) Their pretest scores were not significantly different than the other groups; (2) Their socioeconomic context as indicated by the percent of students that qualify for FRL was approximately equal, if not slightly at the lower range, of the other implementing groups; and (3) The teachers for this group were also not better prepared (i.e., similar rates of master's or undergraduates degrees in biology).

It is important to emphasize that similar gains without PD would be unlikely if the curriculum materials were not educative, based on prior research findings (Cervetti et al., 2015; Davis et al., 2014; Stein & Kaufman, 2010). The educative curriculum materials provided to implementing teachers here were of high quality, meeting almost all of the criteria put forth by Beyer et al. (2009). This demonstration of student gains in qualitative predictions across PD conditions suggests that, at least in some situations, high-quality educative curriculum materials may be sufficient to permit certain student learning gains with novel learning approaches. This finding counters the hypothesis that educative curriculum materials may not be sufficient to promote student learning gains (Davis & Krajcik, 2005; Schneider & Krajcik, 2002). Of course, it remains an open question whether there were also other critical features within this particular unit that led to robust student gains. For example, the unit was designed so that teachers and students were asked to revisit specific core ideas in biology during the course of the unit (the scientific mechanisms of meiosis and fertilization) in multiple contexts (Schuchardt & Schunn, 2016). Therefore, learning may be more robust to variations in PD, because it did not critically depend on successful teacher implementation for any single activity.

All implementing conditions showed some student learning gains in quantitative predictions. However, when time spent on this predictive practice in face-to-face PD was decreased from 23 to 8 hours, student learning gains were significantly lower, but still comparable to learning gains of students in classrooms in which traditional instruction in inheritance occurred (reported in Schuchardt & Schunn, 2016). This result contrasted with that obtained for qualitative predictions that showed essentially no change when face-to-face PD was decreased from 23 to 8 hours. The difference in the effect of decreasing face-to-face PD on the two predictive practices points to the importance of considering intended student learning outcomes when making decisions about investments in PD. Even when the practice (making predictions) is apparently the same, because qualitative and quantitative predictions require students to draw on different resources, student learning outcomes may be affected differently by changing the amount of support that teachers receive.

When PD was dropped from 8 to 0 hours, no significant further reduction was observed in students' ability to make quantitative predictions. Consideration of threshold effects is important in optimally deploying school district resources (Archibald, Coggshall, Croft, & Goe, 2011); although it may be counterintuitive to some administrators, here we have an example in which providing a small amount of PD to teachers in this difficult content area had no benefit over providing only access to educative materials. This result combined with that on qualitative predictions suggests that in terms of scalability, once the decision has been made to reduce PD, eliminating PD if teachers are provided with educative curriculum materials may do little additional harm.

At a minimum, these results provide support for the idea that when designing and assessing PD, there is a need to move beyond (1) general questions such as whether the presence of PD has an effect or (2) specifying fixed general guidelines regarding the amount of PD. Instead, we need to ask more subtle questions that explore what kind of student learning outcomes requires what kinds and amounts of teacher support (Borko, 2004; Dede et al., 2009). We would argue that considering how much PD to offer is a multifaceted problem that involves (1) Intended student learning outcomes, (2) prior teacher preparation, (3) topics covered in PD and educative curriculum materials, and (4) consideration of threshold effects. The intended student learning outcomes in this study (quantitative predictions and qualitative predictions) are areas in which students receiving traditional instruction in inheritance have been shown to struggle (Schuchardt & Schunn, 2016; Stewart, 1983). However, teachers are much more comfortable with and have received more extensive training in the conceptual underpinnings of inheritance (the biological mechanisms) that underlay both types of predictions, and are less comfortable with and have received less training in the mathematical concepts associated with the quantitative predictions (Cox et al., 2016; Furner & Kumar, 2007; Offer & Mireles, 2009; Sorgo, 2010). Thus, it is reasonable to suppose that to achieve the same effect of PD on student learning in qualitative and quantitative predictions, teachers may need more time spent on the quantitative underpinnings. In the 23Hour PD condition, content-focused instruction occurred with teachers working through the unit as students with one exception: Three hours spent on having teachers decipher and address common student errors in making quantitative predictions. This and an additional 5 hours of content-focused PD were eliminated when PD was reduced to 8 hours. All of the content-focused PD (including addressing student errors in making quantitative predictions) involved intertwined discussions of biological and quantitative conceptual understandings. Therefore, we could not determine which aspect of the content focused PD that was removed from the 23Hour PD condition was responsible for the result that quantitative prediction gains decreased while qualitative prediction gains remained the same. In other words, we are unable to determine whether the observed student learning declines in quantitative predictions were caused by the decrease in time teachers spent grappling with the unit as students, or the elimination of the PD task of deciphering and addressing student errors in making quantitative predictions. Given the experimental design, we are also unable to address the specific effects of removing a comparable amount of pedagogically focused PD and whether that was responsible for the differential effects on student learning outcomes. It seems the question for future research becomes: When considering both intended student learning outcomes and prior teacher preparation, what experiences in face-to-face PD coverage in the context of educative curriculum materials will provide the maximal benefit in the least amount of time.

4.1 | Limitations and future directions

This is a quasi-experimental study conducted in a real-world context. As such, it has the benefit of real-world applicability, but trade-offs needed to be made when assigning teachers to conditions to balance administrative demands, participating teacher investment and needs, and experimental considerations. For example, the conditions had different numbers of participating teachers because of the need to balance administrative demands and participating teacher investment and needs. However, every effort was made to balance conditions on dimensions known to contribute to learning such as teacher education and experience and percent of students qualifying for FRL. Statistical analyses revealed no significant differences across conditions on these variables or on students' pretest performance on the learning outcomes.

While this study included at least 160 students within each condition, the number of teachers within each condition ranged from four to eight. The small numbers of teachers within each condition leaves the study open to questions about generalizability. Mitigating against this objection are a few factors. First, preexisting differences that are often associated with student learning outcomes (preexisting scores and economic circumstances) were accounted for in the statistical analyses. Second, including as a covariate a measure of student/teacher learning known to be unaffected by the intervention (as an indicator of generalized learning) had no effect on the findings reported here. Third, when a measure often associated with improved student performance in science, teacher science education level (Monk, 1991), was assessed as the independent variable instead of PD time, no significant differences were found associated with teacher science education level. Fourth, differences in condition means did not appear to be caused by outliers in teacher means.

In this study, which arose out of a larger study on learning gains in science from a curriculum centered on mathematical modeling and aligned with NGSS, we separate out the effects of PD on two different aspects of the scientific practice of making predictions. The chain of causality from teacher PD to student learning outcomes is complex, likely including various changes in teacher content knowledge and instructional practices during unit enactment, and then changes in student enactment of activities (Desimone, 2009; Kleickmann et al., 2016; Stein, Grover, & Henningsen, 1996). These are not addressed in this study. However, research that only studies these intervening steps and never considers student learning is incomplete, and a research effort that studies how PD changes teachers, documents the changes in teaching, and examines the effects on students would necessitate a very large investment of resources (Luft & Hewson, 2014). Before such large-scale studies are conducted, it has been recommended that initial studies are needed to understand how variations in PD affect different aspects of student learning (Luft & Hewson, 2014). We show here that it is possible in the context of educative curriculum materials to see differential effects of time invested in PD on an assessment where the practice (making predictions) is the same, but the resources students draw on (applying conceptual understanding of the mechanism, vs. solving a quantitative problem to produce a numerical answer) are different. Thus, we address an oft-cited but inadequately supported claim that teachers will need differential support for implementing different aspects of NGSS-aligned curricula (National Research Council, 2015; Wilson, 2013).

In general, the exact reasons for these science practice specific effects of face-to-face PD support in the context of implementation of an NGSS-aligned unit when teachers are provided with educative curriculum materials is unclear based on this study. However, the results of this study suggest that when making decisions about scalability of NGSS and investment of resources into PD, it is necessary to consider the desired specific student learning outcomes within a unit. Such considerations allow for efficient deployment of PD resources, which is critical in scaling units to all the contexts in need of new rigorous science curriculum materials (Archibald et al., 2011; Bybee, 2014; Reiser, 2013). Furthermore, in terms of research, the results presented here suggest productive avenues concerning the interaction of teacher support requirements for learning outcomes, and how they can be met. As such, it seems necessary to reiterate the call for more small-scale studies on PD that include research on student learning outcomes (Luft & Hewson, 2014) to guide investment on the necessary and larger-scale studies that focus on connections between student outcomes and teachers' instructional practices, cognitions, and beliefs (Desimone, 2009; Luft & Hewson, 2014; van Driel, Merink, van Veen, & Zwart, 2012; Yoon et al., 2007).

ENDNOTE

[1] A similar pattern is found if the analysis is conducted on the difference of mean posttest and pretest scores by predictive practice for each teacher.

REFERENCES

Archibald, S., Coggshall, J. G., Croft, A., & Goe, L. (2011). High-quality professional development for all teachers: Effectively allocating resources. Research & policy brief. *National Comprehensive Center for Teacher Quality*.

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

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is: Or what might be: The role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–8.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in history and philosophy of biological and biomedical sciences*, *36*, 421–441.

Becker, K., & Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education*, *12*(5-6), 23–37.

Berlin, D. F., & Lee, H. (2005). Integrating science and mathematics education: Historical analysis. *School Science and Mathematics*, *105*(1), 15–24.

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

Bing, T. J., & Redish, E. F. (2008). Symbolic manipulators affect mathematical mindsets. *American Journal of Physics*, *76*, 418–424.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of Science Teacher Education*, *25*, 211–221.

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

Cox, C., Reynolds, B., Schuchardt, A., & Schunn, C. (2016). How do secondary level biology teachers make sense of using mathematics in design-based lessons about a biological process? In *Connecting science and engineering education practices in meaningful ways* (pp. 339–371). Springer International Publishing.

Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, *78*(10), 1056–1063.

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

Davis, E. A., Palinscar, A. S., Arias, A. M., Bismack, A. S., Marulis, L. M., & Iwashyna, S. K. (2014). Designing educative curricula materials: A theoretically and empirically driven process. *Harvard Educational Review*, *84*(1), 24–52.

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, *60*(1), 8–19.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. *Journal of Research in Science Teaching*, *51*(5), 635–638.

Doppelt, Y., Schunn, C. D., Silk, E. M., Mehalik, M. M., Reynolds, B., & Ward, E. (2009). Evaluating the impact of a facilitated learning community approach to professional development on teacher practice and student achievement. *Research in Science and Technological Education*, *27*(3), 339–354.

Furner, J. M., & Kumar, D. D. (2007). The mathematics and science integration argument: A stand for teacher education. *Eurasian Journal of Mathematics, Science & Technology Education*, *3*(3), 185–189.

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

Gupta, A., & Elby, A. (2011). Beyond epistemological deficits: Dynamic explanations of engineering students' difficulties with mathematical sense-making. *International Journal of Science Education*, *33*(18), 2463–2488.

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.

Hurley, M. M. (2001). Reviewing integrated science and mathematics: The search for evidence and definitions from new perspectives. *School Science and Mathematics*, *101*(5), 259–268.

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.

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

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

Lee, M. M., Chauvot, J. B., Vowell, J., Culpepper, S. M., & Plankis, B. J. (2013). Stepping into iSMART: Understanding science-mathematics integration for middle school science and mathematics teachers. *School Science and Mathematics*, *113*(4), 159–169.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, *41*(3), 635–679.

Levy, S. T., & Wilensky, U. (2009). Crossing levels and representations: The Connected Chemistry (CCI) Curriculum. *Journal of Science Education and Technology*, *18*(3), 224–242.

Lin, S.-F., Lieu, S.-C., Chen, S., Huang, M.-T., & Chang, W.-H. (2012). Affording explicit-reflective science teaching by using an educative teachers' guide. *International Journal of Science Education*, *34*(7), 999–1026.

Luft, J. A., & Hewson, P. W. (2014). Research on teacher professional development programs in science. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research in science education* (pp. 889–909).

Lyons, K. C. (2013). *2012 National Survey of Science and Mathematics Education: Status of high school biology*. Chapel Hill, NC: Horizon Research, Inc.

Malone, K. L. (2008). Correlation among knowledge structures, force concept inventory, and problem-solving behaviors. *Physical Review Special Topics Physics Education Research*, *4*, 1–15.

Marek, E. A., & 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.

Monk, D. H. (1991). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125–145.

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

National Research Council. (2015). *Guide to implementing the Next Generation Science Standards*. Washington, DC: The National Academies Press.

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

Offer, J., & Mireles, S. V. (2009). Mix it up: Teachers' beliefs on mixing mathematics and science. *School Science and Mathematics*, *109*(3), 146–152.

Radford, D. L. (1998). Transferring theory into practice: A model for professional development for science education reform. *Journal of Research in Science Teaching*, *35*, 73–88.

Reiser, B. J. (2013). *What professional development strategies are needed for successful implementation of the Next Generation Science Standards*. Paper presented at the Invitational Research Symposium on Science Assessment, The Center for K–12 Assessment & Performance Management at ETS.

Roth, W.-M., & Bowen, G. M. (1994). Mathematization of experience in a grade 8 open-inquiry environment: An introduction to the representational practices of science. *Journal of Research in Science Teaching*, *31*(3), 293–318.

Roth, W.-M., Tobin, K., & Shaw, K. (1997). Cascades of inscriptions and the re-presentation of nature: How numbers, tables, graphs and money come to re-present a rolling ball. *International Journal of Science Education*, *19*(9), 1075–1091.

Scher, L., & O'Reilly, F. (2009). Professional development for K–12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness*, *2*, 209–249.

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

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

Schuchardt, A. (2016). *Learning biology through connecting mathematics to scientific mechanisms: Student outcomes and teacher supports* PhD Doctoral Dissertation. Pittsburgh, PA: University of Pittsburgh. 10298845.

Schuchardt, A., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Science Education*, *100*, 290–320.

Sorgo, A. (2010). Connecting biology and mathematics: First prepare the teachers. *CBE-Life Sciences Education*, *9*, 196–200.

Spillane, J. P., Gomez, L., & Mesler, L. (2009). Notes on reframing the role of organizations in policy implementation. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 409–425). Washington, DC: American Educational Research Association.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

Stein, M. K., & Kaufman, J. H. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal*, *47*(3), 663–693.

Stewart, J. (1983). Student problem solving in high school genetics. *Science Education*, *67*(4), 523–540.

Taasoobshirazi, G., & Glynn, S. M. (2009). College students solving chemistry problems: A theoretical model of expertise. *Journal of Research in Science Teaching*, *46*(10), 1070–1089.

Tang, K.-S., Tan, S. C., & Yeo, J. (2011). Students' multimodal construction of the work-energy concept. *International Journal of Science Education*, *33*(13), 1775–1804.

Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A. M., & Tipton, E. (2016). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness*, *10*(2), 1–31.

Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics Physics Education Research*, *3*, 1–22.

van Driel, J. H., Merink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education*, *48*(2), 129–160.

Watanabe, T., & Huntley, M. A. (1998). Connecting mathematics and science in undergraduate teacher education programs: Faculty voices from the Maryland Collaborative for Teacher Preparation. *Connecting Mathematics and Science*, *98*(1), 19–25.

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A modeling method for high school physics instruction. *American Journal of Physics*, *63*(7), 606–619.

Wilson, S. M. (2013). Professional development for science teachers. *Science*, *340*, 310–313.

Wu, H.-K., & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms: A case study of seventh graders' use of data tables and graphs. *Journal of Research in Science Teaching*, *43*(1), 63–95.

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. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1)*.

How to cite this article: Schuchardt AM, Tekkumru-Kisa M, Schunn CD, Stein MK, Reynolds B. How much professional development is needed with educative curriculum materials? It depends upon the intended student learning outcomes. *Sci Ed*. 2017;101:1015–1033. https://doi.org/10.1002/sce.21302