Journal of Science Education and Technology
https://doi.org/10.1007/s10956-018-9756-y

Measuring Pedagogy and the Integration of Engineering Design in STEM Classrooms

Tory Williams[1] · Jonathan Singer[1] · Jacqueline Krikorian[1] · Christopher Rakes[1] · Julia Ross[2]

Abstract
The present study examined changes in high school biology and technology education pedagogy during the first year of a three-year professional development (PD) program using the INSPIRES educative curriculum. The Next Generation Science Standards (NGSS) calls for the integration of science and engineering through inquiry-based pedagogy that shifts the burden of thinking from the teacher to the student. This call is especially challenging for teachers untrained in inquiry teaching and engineering or science concepts. The INSPIRES educative curriculum materials and PD provided a mechanism for teachers to transform their teaching to meet the NGSS challenges. This study followed a longitudinal triangulation mixed methods design. Selected lessons were video recorded, scored on the Reformed Teaching Observation Protocol (RTOP) rubric, and examined for qualitative trends. Year 1 results indicated that teachers had begun to transform their teaching and pointed to particular lessons within the INSPIRES curriculum that most facilitated the reform. Instructional practices of participants improved significantly as a result of the INSPIRES PD program and also aligned with previous, similar studies. These findings provide insights for rethinking the structure of professional development, particularly in the integrated use of an educative curriculum aligned with intended professional development goals.

Keywords Educative curriculum · Engineering education · Mixed methods · Pedagogical reform · Professional development

Introduction

The publication of the *Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council 2012) and the subsequent adoption of Next Generation Science Standards (NGSS) have led to a significant shift in instruction and student learning expectations in K–12 science classrooms (Cuban 2013; Roseman et al. 2015). In addition to the use of student performance expectations, the NGSS has multiple components that are significantly different from past reforms, including the incorporation of science and engineering "practices," "disciplinary core ideas," and "crosscutting concepts" (Next Generation Science Standards 2013). These changes to STEM teaching and learning will require both the need for new curricular materials, as well as support in reformed instructional practices (Richmond et al. 2016; Fishman et al. 2017; Ross et al. 2015; Singer et al. 2016). For example, inclusion of pedagogical practices, such as coaching student groups through an open-ended design challenge and probing students for science- or math-based rationale, supports success in addressing the NGSS. Teacher professional development (PD) is a critical strategy for supporting in-service educators in the use of new materials and the implementation of reform-based instructional practices (Reiser 2014). This shift presents significant challenges to teachers unfamiliar with engineering-based pedagogy and engineering or science concepts.

The INcreasing Student Participation, Interest, and Recruitment in Engineering and Science (INSPIRES)

✉ Tory Williams
twilliams@umbc.edu

Jonathan Singer
jsinger@umbc.edu

Jacqueline Krikorian
jkrikor1@umbc.edu

Christopher Rakes
rakes@umbc.edu

Julia Ross
rjulie@vt.edu

[1] Department of Education, University of Maryland Baltimore County, Baltimore, MD, USA

[2] College of Engineering, Virginia Tech, 620 Drillfield Drive, Blacksburg, VA 24061, USA

Published online: 03 November 2018

curriculum is written for grades 9–12 and focuses on integrating all areas of STEM. These materials use a real-world engineering design challenge (building a functional hemodialysis system for an adolescent patient) and inquiry-based learning strategies (e.g., phenomena-first, artifact sharing, probing questions) to engage students, increase technological literacy, and develop key practices foundational for success in STEM disciplines. The curriculum was designed to be flexible, low cost, and approximately three weeks in length (Ross et al. 2015). The curriculum is well-aligned to the ideas and practices of engineering articulated in the *Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council 2012). As a result, the INSPIRES curriculum targets all four NGSS Engineering Design performance expectations (HS-ETS1) and all eight Science and Engineering Practices (Next Generation Science Standards 2013). In addition, the INSPIRES curriculum has been constructed to include explicit, imbedded supports that highlight specific elements in the lesson plan that may impact student learning. The inclusion of these elements may support teachers "to learn about teaching within the curriculum materials, making them educative" (Schneider et al. 2005). The educative curriculum materials support teachers by including features that encourage reflection and promote connections among specific content, pedagogy and pedagogical-content knowledge (Ball and Cohen 1996; Schneider et al. 2005; Knaggs and Schneider 2012). These characteristics make INSPIRES unique compared to other currently available engineering-based curriculum materials (Ross et al. 2015). Within each INSPIRES lesson, the educative components appear in a column adjacent to particular sections that are potentially challenging for teachers or learners. Similar to the support described by Davis and Krajcik (2005), the INSPIRES educative traits highlight strategies or information that is intended to address (among other things) student misconceptions, additional content knowledge for teachers, potential probing questions, or specific pedagogical strategies. For example, in INSPIRES lesson 7, *Introduction to Dialysis*, the lesson plan describes how the teacher can facilitate student experiments that explore the movement of "waste" products across a semi-permeable membrane. Here, the educative elements include (1) highlighting student misconceptions related to "equilibria," (2) teacher content knowledge regarding experimental variables that impact the "rate of diffusion" versus the amount of "mass transfer," and (3) a description of how the lesson moves from a macroscopic phenomenon to a particle-level simulation.

The present study explored the benefits and limitations of infusing the INSPIRES educative curriculum materials within a professional development (PD) system. Such an enhancement of PD is posed as a mechanism for strengthening teacher pedagogical skills for integrating engineering practices in high school biology and technology education classrooms. The research questions were the following:

1) Did teachers' classroom practice change as a function of INSPIRES-based professional development and curriculum enactment as measured by the Reformed Teaching Observation Protocol (RTOP)?
2) Did teacher pedagogical skill development differ for biology and technology education teachers?

Conceptual Framework

The Professional Development: Research, Implementation, and Evaluation (PrimeD) framework (Saderholm et al. 2016) guided the PD throughout the study. Elements of the PrimeD framework were developed through a synthesis of PD theory from multiple sources, such as Darling-Hammond and McLaughlin (1995), McAleer (2008), Desimone (2009), Loucks-Horsley et al. (2010), and Sztajn (2011). PrimeD divides PD into the following four phases: design and development, implementation, evaluation, and research. In the design and development phase, the PD providers met with district personnel and teachers to develop a common vision and design, including the establishment of goals, strategies, needs assessment, targets, and contextual factors (*challenge space*). The implementation phase consisted of cycles of whole and small group meetings and utilized classroom implementation activities. Whole group meetings occurred during summer workshops and periodically throughout the school year. Small group meetings occurred during the school year between whole group meetings. Classroom implementation activities were guided by Plan-Do-Study-Act (PDSA) cycles (Bryk et al. 2011). For each PDSA cycle, teachers implemented activities to address a particular challenge discussed during a whole or small group meeting. Teachers collected artifacts during classroom implementation to bring back to the whole and small group meetings. Feedback was provided throughout each phase of the program and findings initiated a revisiting of the challenge space prior to subsequent rounds of implementation. Research goals, design, data, threats to validity and reliability, and ongoing results were an integral component of the development and adjustment of the challenge space. However, even with effective PD programs, research has shown that teachers struggle to successfully integrate engineering design- and inquiry-based practices (Schneider et al. 2005).

Educative Curriculum

The integration of educative curriculum materials with PD has shown promise in small-scale studies (e.g., Rushton et al.

2011; Singer et al. 2011; Lotter et al. 2013). In a PD guided by an educative curriculum, the curriculum acts as a scaffold to illustrate pedagogical principles to be transferred to teaching practice. In this study, the classroom enactment of the educative materials (INSPIRES) was intended to be a critical component of the PD strategy. Thus, teachers were given the guided experience of grappling with the educative materials from both the student's and teacher's perspectives, followed by reflective discussions on the lessons' pedagogical design. These experiences provided opportunities for teachers to encounter the affordances and limitations of each activity from the student's perspective and then to discuss the rationale for how the activity was constructed and how it may be adapted (Remillard 2000). The curricular materials serve as a scaffold by providing the teachers concrete examples for how to translate abstract ideas into a tangible useful product. Employing such a strategy may promote significant change in the content knowledge and pedagogical practices of high school STEM teachers (Singer et al. 2011; Lotter et al. 2013). Arias et al. (2016) found that teachers better supported students in qualifying predictions, forming evidence-based claims, documenting observations, and planning next steps when utilizing an educative curriculum for electric circuits; educative features included practice overviews, in-lesson "how and why supports," practice reminder boxes, rubrics, examples, and narratives. Teachers have reported that enacting the educative curriculum profoundly changed their attitudes and methods for teaching science (Pringle et al. 2017). With the proper educative features, these curricula are already thought to be appropriate for addressing challenges of the NGSS (Roseman et al. 2017). Additionally, there is a call to further shift teachers' perspective of educative materials from merely a source of student activities to a dynamic tool for supporting teachers' own pedagogical growth (McNeill et al. 2017).

INSPIRES Educative Curriculum and PD Program

The INSPIRES educative curriculum materials and accompanying teacher PD framework are intended to facilitate teacher adoption of design-based pedagogical practices necessary for integrating engineering and biology concepts and practices. The PD program began with a 5-day summer institute (SI) followed by a series of 2-h, monthly sessions sustained across the academic year. The year 1 SI focused on the following four key components: (1) the INSPIRES educative curriculum materials, (2) STEM practices, (3) pedagogical practices, and (4) reflective critiques. The INSPIRES hemodialysis materials were developed to model and scaffold the other three components. During the STEM practices segment of summer PD, specific activities from the pre-selected materials were used by the facilitators to illustrate key ideas or as "jumping off" points for deeper discussion. The key foci of the STEM practices component were on building content knowledge, an understanding of the engineering design process, and skills with the tools needed for the design challenge. Teacher teams participated in the curriculum as students and performed all design-, build-, and test-based engineering activities. The key focus of the pedagogical practices component was on building pedagogical content knowledge. Core elements of this component focused on modeling various pedagogical strategies, STEM practices, and curriculum materials. Example practices that were emphasized include phenomena-first, inquiry, and design-based learning (e.g., Predict, Observe, Explain; integration of an engineering design loop), collaboration (e.g., jigsaws, Think-Pair-Share), context (e.g., driving questions, KWL charts), technology integration (e.g., simulations, data collection), and sense-making and assessment (e.g., wait time, probing questions, prior knowledge). The reflective critiques component supported both STEM and pedagogical practices as well as classroom management issues. Following each lesson, the PD facilitators engaged teachers in discussions relating the lessons' content to its structure and strategies.

Method

The INSPIRES Curriculum

The INSPIRES curriculum was developed to integrate engineering design principles into high school science and technology classes. The present study used *Engineering in Health Care: Hemodialysis*, one of five modules that comprise the INSPIRES curriculum (Ross et al. 2015). In this module, students learn about kidney function, dialysis, diffusion of waste across membranes, and factors that influence mass transfer and diffusion rates. By the end of the module, students design, build, test, and revise an apparatus that mimics the function of a hemodialysis system. The module applies a project-based approach (Blumenfeld et al. 1991; Marx et al. 2004; Krajcik and Blumenfeld 2006; Willis 2018), in which the design challenge is introduced at the beginning of the module and is used throughout multiple lessons to drive the learning of important science and engineering concepts.

Participants

The present study was conducted in collaboration with a large mid-Atlantic public school system. With 174 schools, programs, and centers, nearly 9000 classroom teachers and over 105,000 students, this district is one of the largest school systems in the US. The district's 800,000+ residents live in suburban, rural, and urban neighborhoods comprising of cultures and backgrounds representative of the nation's diversity. Overall, 54.8% of the district's students represent racial and ethnic groups other than White, 48.9% are female, and 44.8% are eligible for free/reduced price meals.

Twenty-seven biology and technology education teachers from eleven high schools participated in the study. These schools represent traditional and alternative schools that offer both biology and technology education courses and form a representative cross-section of the district. The group of teachers included both males ($N = 16$) and females ($N = 11$) who reported their race/ethnicity as Black or African American (22%) or White (78%) and whose classroom teaching experience ranged from 2 to 28 years (16% of teachers had 0–5, 47% had 6–10, 26% had 11–15, 11% had > 15 years of experience).

Data Sources

The data presented in this study represents those from the first year of a three-year, longitudinal research project. The data were obtained from scoring classroom videos at four time points. The first data point (Baseline lesson) was collected during the spring prior to the summer PD event in which the teachers were asked to provide their best attempt of incorporating NGSS Engineering Design Standards (HS-ETS1) into a lesson. This same prompt was utilized approximately 1 year later during the following semester to serve as a measure of potential growth during year 1 (Transfer lesson). Two additional lessons associated with the enactment of the INSPIRES educative materials were recorded during the intervening fall (Lessons 7 and 11).

The INSPIRES Hemodialysis Lesson 7 is structured as a phenomena-first, science-rich, inquiry activity. It provides an opportunity for students to collect visual and quantitative evidence of "waste" removal from artificial blood by diffusion. The lesson's base activity involves dialysis tubing formed into a "bag" and filled with 20 ml of simulated blood. The dialysis bag is then placed in a beaker of water. By identifying and altering variables (e.g., porosity of the bag membrane, water temperature), the conditions affecting waste removal and, therefore, diffusion, can be identified and tested. This creates opportunities for students to work collaboratively in teams, identify experimental variables, form predictions, design protocols and procedures, and carry out experiments. In addition, the lesson is designed to allow student teams to share results with the whole class, analyze data, and reflect on outcomes. The strategy of sharing results is expected to deepen understanding of the critical scientific concepts and to inform design choices in the larger design challenge.

The objective for INSPIRES Lesson 11 was for students to apply the knowledge and experiences from all previous INSPIRES lessons and use the design process to design, build, and test a hemodialysis system. Lesson 11 begins with a review of the design challenge, the various preceding activities, and the connections between activities that address the challenge. Teams are shown various supplies (e.g., tubes, membranes, pumps, bottles) and are prompted to plan their designs. Before construction can begin, the teacher probes teams for evidence-based rationale for their various design decisions. Research-based observations of Lesson 11 typically captured the design phase and sometimes the beginning of the build phase. Overall, Lesson 11 was crafted to lapse 2–3 class periods where students could continue building their systems, complete testing, and further revise their design.

Collected classroom videos were scored using the RTOP observational instrument. The RTOP was developed by the Arizona Collaborative for Excellence in the Preparation of Teachers to capture current elements of pedagogical reform. The instrument was written based on constructivist theory and with national standards of math and science in mind. The RTOP is widely applied in STEM educational research as both a quantitative and qualitative tool (e.g., MacIsaac et al. 2001; Enderle et al. 2014; Amolins et al. 2015) by outlining characteristics of reform in a 25-item rubric on a 0–4 performance scale. The training manual defines level 0 as "not descriptive of the lesson" and level 4 as "very descriptive of the lesson," and prior psychometrics on the RTOP instrument revealed an "exceptionally high" estimate of reliability (Piburn and Sawada 2000).

RTOP items are divided into the following five subcategories: Lesson Design, Propositional Knowledge, Procedural Knowledge, Classroom Culture, and Teacher–Student Relationships. Lesson Design items ask the extent to which class instruction incorporates prior knowledge, social construction of knowledge, the progression from concrete to abstract concepts, valuing multiple solutions or approaches, and flexibly in following students' ideas or needs. Items in the Propositional Knowledge subcategory ask whether significant STEM ideas are the focus, if explicit connections are made between STEM ideas and with real-world applications, and the extent of teacher comfort and expertise in the STEM content. Rating Procedural Knowledge items will indicate the extent of multiple means of representation and the opportunity for students to make predictions, think critically, reflect on learning, and engage in argumentation. Items representing Classroom Culture assess multiple means of expression, the facilitation of divergent thinking, the value of student discourse, and the classroom as a safe place to express individual ideas. Finally, Teacher–Student Relationship items evaluate the level of leadership and empowerment passed from teacher to students, intended use of wait time, and teacher facilitation of student understanding (Piburn and Sawada 2000).

Prior to data collection, four coders were trained to identify the characteristics of each RTOP item and performance level. The coders developed and refined performance indicators within the RTOP rubric to bring validity to particular score levels and to enhance inter-rater reliability. Classroom video data were deidentified by replacing teacher names with random numeric codes. Subjectivity was further discouraged by frequent checks of inter-rater reliability; the four coders achieved high agreement despite their varied expertise within STEM fields or education. Twenty percent of the videos was coded by all four researchers with an additional 14% being double coded. Interclass correlation coefficients (K) that ranked in the range of 0.75–1.00 were considered excellent, and ranks between 0.60 and 0.74 were considered good (Cicchetti 1994). Interclass correlation coefficients for videos scored by all four coders were the following: Baseline lesson ($K = 0.705$), Lesson 7 ($K = 0.826$), Lesson 11 ($K = 0.711$), and Transfer lesson ($K = 0.718$). For all co-scored videos, discrepancies in item scores between raters were deliberated on until mutual consensus was reached. Classroom videos were given a performance level score (0–4) on all 25 items in the RTOP rubric. Summing scores within each subcategory and then averaging across all teachers yielded summary performance within subcategories. Summing scores of all 25 items and then averaging across teachers determined summary total RTOP performance.

Data Analyses

For statistical analysis, each teacher's video received a single score for each subcategory by averaging the scores for its five items. Overall trends were identified during the first year of the study by relating teacher instruction of the four lessons (Baseline lesson, Lesson 7, Lesson 11, and Transfer lesson). Additionally, a total average score was computed for all 25 RTOP items. Differences in total and subcategory averages across the four lessons were analyzed with a repeated measures analysis of variance (ANOVA) with one fixed factor to compare biology and technology education teachers.

A subsample of teacher participants was selected for qualitative analysis. Raters further characterized typical practices that were generally representative of qualitative traits observed in Baseline and Transfer lessons. The systematic approach used in this characterization involved the selection of three biology and three technology education teachers whose Baseline RTOP scores were in the mean range for at least two out of three of the following subcategories: Procedural Knowledge, Classroom Culture, or Teacher–Student Relationships. Focus was placed on these subcategories as they represented areas of notable growth between Baseline and Transfer lessons for teachers overall. By selecting teachers whose assigned RTOP scores were around the means representative to all teachers, the raters aimed to capture the common traits of teaching practices at the different time points of the study. Further, each focal teacher represented a different high school in the district. This systematic approach was adapted from both domain analysis methods (Spradley 1980) and analytic coding techniques (Coffey and Atkinson 1996).

Raters critically examined the RTOP scoring notes and lesson summaries for focal teachers' lessons across the four time points. For each lesson, the raters reached consensus on identifiable pedagogical traits. Themes were recognized across all six focal teachers' Baseline lessons which led to the development of a typical Baseline lesson qualitative description. The process was repeated respective to Lesson 7, Lesson 11, and the Transfer lessons.

Results

Quantitative Analysis

Total RTOP scores were averaged for all 25 items and for each subcategory (Table 1). At the Baseline, teachers scored an average of about half the possible points, indicating that they were not initially teaching with strong reform pedagogies. Lesson 7 scores were similar to Baseline scores. For Lesson 11, teachers scored approximately two thirds of the possible points. The Transfer lesson scores were slightly lower than those of Lesson 11.

The repeated measures ANOVA indicated significant differences across the four lessons for the overall RTOP as well as for all subcategories except for Propositional Knowledge (Table 2). No significant differences were found between the biology and technology education teachers (Table 2).

Pairwise comparisons revealed that the Baseline and Lesson 7 scores were not significantly different except in the case of the Procedural Knowledge subcategory. Lesson 11, however, scored significantly higher than all other lessons for the overall RTOP for all teachers (Table 3). Further, Transfer lessons scored significantly higher than Baseline lessons for both the Classroom Culture and Teacher–Student Relationships subcategories.

Qualitative Analysis

Here, we evaluate qualitative themes that reflect shifts in teacher instruction across the four focal lessons. Evidence of qualitative trends fits into the following six themes: Guided vs. Open Strategy, Probing of Prior Knowledge, Making Predictions, Making Connections, Student Reflection, and Teacher Sharing (Tables 4–7). Guided vs. Open Strategy highlights traits that may characterize a lesson as either more prescribed or open-ended. Probing of Prior Knowledge characterizes the degree to which teachers facilitate students' application of prior knowledge to the current lesson. Making Predictions refers to elements of prediction formulation, justification, and verification that may occur throughout a STEM lesson. Making Connections highlights instances where teachers or students explicitly think about how past lessons inform the current lesson or how the current lesson may inform future lessons. Student Reflection captures elements of divergent and critical thinking and the strategies used to

Table 1 Mean total scores for RTOP overall and subcategories

RTOP categories	Baseline Mean total score (SD)	Lesson 7	Lesson 11	Transfer
All teachers (*N* = 27)				
Overall[a]	50.7 (12.0)	56.3 (11.5)	68.5 (10.4)	59.6 (11.1)
Lesson design[b]	9.2 (3.8)	10.5 (3.1)	14.2 (2.4)	11.4 (3.9)
Propositional knowledge[b]	13.9 (2.3)	14.4 (3.2)	14.4 (3.2)	14.1 (2.4
Procedural knowledge[b]	9.0 (2.9)	10.9 (1.9)	13.8 (1.8)	11.0 (2.6)
Classroom culture[b]	8.6 (2.7)	9.8 (2.1)	12.7 (2.4)	11.3 (2.4)
Teacher–student relationships[b]	10.0 (2.9)	10.8 (2.7)	13.4 (2.1)	11.9 (2.4)
Biology teachers (*N* = 14)				
Overall[a]	50.4 (10.6)	59.5 (8.5)	70.1 (7.2)	57.4 (10.6)
Lesson design[b]	9.0 (3.3)	10.8 (2.5)	14.4 (1.8)	10.7 (3.8)
Propositional knowledge[b]	14.4 (2.3)	15.4 (2.7)	14.8 (2.3)	14.1 (2.1)
Procedural knowledge[b]	8.9 (2.7)	11.5 (1.9)	14.1 (1.3)	10.1 (3.0)
Classroom culture[b]	8.4 (2.2)	10.2 (1.5)	12.8 (2.1)	10.9 (2.6)
Teacher–student relationships[b]	9.8 (2.2)	11.6 (2.5)	14.1 (1.4)	11.6 (2.3)
Technology education teachers (*N* = 13)				
Overall[a]	51.0 (14.0)	52.5 (13.8)	66.6 (13.4)	62.1 (11.6)
Lesson design[b]	9.5 (4.4)	10.1 (3.7)	13.9 (3.0)	12.1 (4.1)
Propositional knowledge[b]	13.2 (2.4)	13.2 (3.8)	14.1 (4.1)	14.1 (2.7)
Procedural knowledge[b]	9.2 (3.2)	10.2 (1.8)	13.5 (2.3)	11.9 (1.9)
Classroom culture[b]	8.8 (3.2)	9.2 (2.6)	12.5 (2.8)	11.8 (2.2)
Teacher–student relationships[b]	10.2 (3.6)	9.8 (2.6)	12.6 (2.5)	12.2 (2.6)

[a] 100 points possible

[b] 20 points possible

support these processes. Teacher Sharing refers to teacher comments that convey personal experiences, notably their struggles while working through the Hemodialysis curriculum as learners. Elements of Teacher Sharing were unique to Lesson 11 (Table 6).

Prior to the first INSPIRES summer PD Institute, the teacher participants were asked to conduct a classroom lesson that addressed their best attempt of incorporating NGSS Engineering Design Standards (HSETS1). This event served as a baseline measure of teachers' initial understanding of integrating engineering design into their instruction. Baseline data revealed that teachers' lessons addressed a wide range of foci varying from classical biological topics, such as evolution and endangered species, to physical sciences, such as propeller designs, fluid flow rates, and simple machines, as well as specialized subjects like forensic science. Despite the large range of topics, multiple themes could be distilled (Table 4).

One emergent Baseline theme was that instruction involved a central activity requiring the collection of data, yet, the activities were confirmational in nature and the introduction of concepts preceded the actual investigation (Table 4, *Guided* vs *Open Strategy*). Additionally, probing for student predictions was limited, and no connections were made between predictions and the corresponding results (Table 4, *Making Predictions*). Baseline lessons typically included connections to prior classroom activities, such as illustrating how the lesson was part of a larger challenge (Table 4, *Making Connections*). While most of the Baseline lessons attempted to make connections to other lessons, limited attempts were

Table 2 Repeated measures ANOVA across four lessons

RTOP categories	Comparisons among Baseline, Lesson 7, Lesson 11, and Transfer *F*(2,50)	Comparisons between biology and technology ed. teachers *F*(2,50)
Total	15.857***	2.067
Lesson design	11.872***	0.819
Propositional knowledge	0.596	1.342
Procedural knowledge	23.667***	2.529
Classroom culture	17.347***	0.766
Teacher–student relationships	12.113***	2.447

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Table 3 Pairwise comparisons for RTOP

RTOP categories	Lesson 7—Baseline Mean difference (SE)	Lesson 11—Baseline	Lesson 11—Lesson 7	Transfer—Baseline	Transfer—Lesson 7	Lesson 11—Transfer
Total	0.250 (0.123)	0.726 (0.098)***	0.476 (0.075)***	0.388 (0.118)*	0.138 (0.118)	0.338 (0.105)*
Lesson design	0.307 (0.192)	1.013 (0.166)***	0.706 (0.113)***	0.472 (0.195)	0.164 (0.202)	0.542 (0.163)*
Propositional knowledge	0.157 (0.145)	0.156 (0.151)	− 0.001 (0.096)	0.070 (0.123)	− 0.086 (0.147)	0.086 (0.159)
Procedural knowledge	0.401 (0.124)*	0.981 (0.102)***	0.580 (0.078)***	0.441 (0.156)	0.041 (0.113)	0.540 (0.116)**
Classroom culture	0.217 (0.126)	0.805 (0.122)***	0.588 (0.091)***	0.557 (0.132)**	0.340 (0.130)	0.248 (0.121)
Teacher–student relationships	0.161 (0.149)	0.674 (0.103)***	0.513 (0.094)***	0.400 (0.135)*	0.239 (0121)	0.274 (0.104)

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

made to integrate student prior knowledge as a means to engage students or adapt the instruction (Table 4, *Probing of Prior Knowledge*). Most of the sampled teachers opened instruction with a traditional drill asking students to provide a definition of a key term related to the day's activity. Generally, student responses were relayed back to the teacher with an emphasis on presenting a correct response.

The INSPIRES Hemodialysis Lesson 7 is structured as a phenomena-first inquiry activity. However, forthcoming qualitative analysis and discussion suggest that the provided, written plan for Lesson 7 was not closely followed by several teachers. Various qualitative traits characteristic of Lesson 7 enactment are listed in Table 5.

In general, Lesson 7 instruction was guided and often teacher-directed (Table 5, *Guided vs. Open Strategy*). Commonly, teachers probed students' prior knowledge of relevant scientific concepts and vocabulary during Lesson 7. Student misconceptions were usually clarified by teachers, but they did not alter the instructional sequence of the lesson (Table 5, *Probing Prior Knowledge*). Teachers typically

Table 4 Qualitative trends among baseline lessons

Theme	Baseline trends
Guided vs. open strategy	More "Hands-on than Minds-on" • Activities are preceded by teacher-centered introduction of key ideas • Teacher provided variables and procedures • Focus on consistent process (doing it correctly) • Activities are used to confirm information presented in the lesson
Probing of prior knowledge	Traditional "Bell work" • Review of prior concepts at start of lesson • Completed as individuals • Ascertained information does not alter instructional sequence
Making predictions	Prediction as "Formality" • Teacher directs individuals to make predictions • Predictions are typically made before the activity
Making connections	Connecting "Past to Present" • Reminds students of introduced concepts from prior lessons • Teacher provides real world examples • Connections mostly "Past to Present"
Student reflection	Traditional "Exit Ticket" • Individuals respond in writing to teacher prompt of student knowledge from the day's lesson • Short, factual information from the day's lesson is the focus of the prompt
Teacher sharing	Not a hallmark of this lesson

Table 7 Qualitative trends among Transfer lessons

Theme	Transfer
Guided vs. open strategy	"Increased Autonomy" • Student groups pursue different approaches • Planning documents utilized • Shallow emphasis on rationale/adaptations
Probing of prior knowledge	"Increased Student Interest" • Presentation of prior results • Base experiment prior to student designing • Increased student engagement
Making predictions	"Shallow Rationale" • Sharing to teacher within groups • Limited pressing for conceptual rationale
Making connections	"Improve Future Results" • Data collected to improve performance or test hypothesis • Mostly implicit connection to concepts discussed prior to investigations • Increased use of "modeling"
Student reflection	"Teacher Prompted Closure" • Limited time set aside at lesson conclusion • Teacher probes and prompts students for key ideas • Superficial response accepted
Teacher sharing	Not a hallmark of this lesson

prompted students to identify possible variables for the experimental system and to make predictions on the effects of changing each variable (Table 5, *Making Predictions*). Making explicit connections to science concepts from a prior lesson was a common practice in enactments of Lesson 7, yet, connections to the engineering design process were sparse (Table 5, *Making Connections*). Student journals were frequently used as a tool to record notes, predictions, data, experimental design plans, and results. Use of notebooks for written reflection on rationale (such as *explaining* the results after experiment completion) was minimal or absent (Table 5, *Student Reflection*).

The objective for INSPIRES Lesson 11 was for students to apply the knowledge and experiences they had acquired from all previous INSPIRES lessons and effectively employ a design process in order to design, build, and test a hemodialysis system. Common qualitative traits are evident across Lesson 11 teacher enactments (Table 6).

During Lesson 11, teachers generally allowed student autonomy by encouraging the development of multiple designs and/or procedures. In addition to following the INSPIRES lesson plan, teachers typically granted students opportunities for divergent thinking by fostering open-ended group work (Table 6, *Guided vs. Open Strategy*). Many teachers facilitated explicit connections to both prior lessons and knowledge (Table 6, *Probing Prior Knowledge*) and established links to the engineering design loop or target (Table 6, *Making Connections*). Students frequently used engineering notebooks for sketching designs or referencing relevant prior knowledge (Table 6, *Making Predictions*, *Student Reflection*). Teachers also referenced their own prior experiences designing, building, and testing hemodialysis systems as they trained in the INSPIRES curriculum (Table 6, *Teacher Sharing*).

For the final class observation of the present study, teachers were asked to select and share a lesson from their repertoire that best highlighted NGSS engineering design practices. Although lesson topics varied widely, the collective group of these lessons is referred to as Transfer lessons. In other words, we wanted to measure how effectively teachers transferred elements of reformed pedagogy, learned through the INSPIRES PD and educative curriculum, into their own original lessons. Table 7 lists common traits evident across teachers' Transfer lessons.

During Transfer lessons, teachers generally allowed some level of student autonomy, demonstrated through students working within small groups and pursuing different approaches to a problem (Table 7, *Guided vs. Open Strategy*). Transfer lessons frequently incorporated strategies to elicit student prior knowledge of a STEM concept (Table 7, *Probing of Prior Knowledge*). Such strategies appeared to spark interest among students and encourage their full participation in the activity. Further, teachers pressed students for shallow levels of rationale, which could be construed as students making predictions about the outcomes of their activity (Table 7, *Making Predictions*). Commonly, students made connections between concepts or multiple activities, during

Table 6 Qualitative trends of Lesson 11 enactment

Theme	Lesson 11 trends
Guided vs. open strategy	Mostly open; "Student Autonomy" • Open-ended group work • Divergent thinking valued through student-determined designs and procedures • Activity has multiple correct solutions • Students encouraged to use additional materials brought from home
Probing of prior knowledge	Relevant "Science Concepts" • Student-selected artifacts or use of KWL charts replaces traditional written drill • Discussion of counter-current flow • Revisiting the relationship between height and flow rate • Cost emphasized over integration of science concepts
Making predictions	Student "Planning" • Design sketching precedes building • Teachers check designs/predictions before students "buy" materials • Groups are expected to combine ideas from multiple designs, or use rationale to select a best design to build
Making connections	Connecting "Past to Present" • Connecting to prior lessons ("Computer Simulation" and "Flow Rate" lessons) • Reminding class of the current step within the engineering design process • References to the multiple criteria and constraints of the design target
Student reflection	Journals used as a "Dynamic Resource" • Notebooks are frequently used for note-taking, data recording, design sketching, and referencing notes from prior lessons to inform design decisions or provide rationale for design decisions
Teacher sharing	"Teachers Share" their own experiences of designing, building, and testing systems • Shared photographs of multiple teacher-built systems • Revealed that teacher systems did not meet all criteria and constraints • Noted that teacher designs were successful without use of pumps

Transfer lessons (Table 7, *Making Connections*). For example, students were engaged in data collection as a means to improve performance or test a hypothesis. Many of the Transfer lessons concluded with a teacher-prompted closure activity that limited the opportunity for student reflection (Table 7, *Student Reflection*). Often, the time reserved for a lesson's conclusion was short in duration, and the discussion was rushed or absent.

Discussion

Quantitative Findings

One striking trend revealed by the quantitative analysis is that enactment of Lesson 11 scored significantly higher than all other lessons overall. As part of the larger INSPIRES Hemodialysis curriculum, Lesson 11 was carefully crafted to incorporate explicit connections to both the engineering design process as well as the underpinning scientific and quantitative rationale. The lesson is further designed to shift the responsibility of learning from the teacher to the students, resulting in a student-centered, inquiry- and project-based experience for learning. Examples of such exemplar lesson traits include but are not limited to the following: small student groups working to communicate designs and procedures, the expectation for science and quantitative rationale to justify design decisions, teachers acting as listeners and facilitators, student groups reporting out and offering critique, teachers enforcing wait time and encouraging divergent thinking, and opportunities to explore phenomena related to real-world engineering challenges (Piburn and Sawada 2000). When Lesson 11 is taught as intended, the resulting RTOP analysis would indicate use of highly reformed pedagogy. Therefore, teachers that made a strong effort to facilitate the lesson as written were well prepared to attain high RTOP scores. Finding high levels of pedagogical reform on this engineering-focused lesson provides support for how quality engineering lessons offer ideal opportunities for student learning. Therefore, teachers equipped with the pedagogical skillset to accompany quality engineering lessons will be better prepared to address the challenges of the NGSS.

Table 5 Qualitative trends of Lesson 7 enactment

Theme	Lesson 7 trends
Guided vs. open strategy	More "Hands-on than Minds-on"
	• Activities preceded by extensive teacher-centered summary of key ideas/vocabulary
	• Student ideas for the activity are solicited; use is limited
	• Variables and procedures are provided by teacher
	• Different groups investigate different variables
	• Teacher discusses results with individual groups
	• Teacher often does calculations of dependent variable for students
Probing of prior knowledge	Traditional "Bell work"
	• Review of prior science concepts at start of lesson
	• Structured as a warm-up (individual student work), followed by class discussion led by the teacher
	• Student prior knowledge does not alter the instructional sequence
Making predictions	Prediction as "Confirmation"
	• Predictions for activities shared within student groups
	• Some teacher probing for information introduced earlier in the lesson as rationale
	• Teacher often confirms prediction rationale before the activity
	• Predictions and rationale discussed mostly prior to the activity
Making connections	"Incomplete" Connections
	• Teacher reminds students of concepts from prior lessons
	• Teacher provides real world examples
	• Superficial connections are made to the engineering design process (e.g., "Where are we?")
	• Connections are mostly "Past to Present"
	• Frequent reference made to reviewing data during the next class
Student reflection	Journals used for "Documentation"
	• Student notebooks used throughout the lesson for notes, predictions, experimental designs, data, results, and to summarize outcomes
Teacher sharing	Not a hallmark of this lesson

Subcategorical RTOP performance revealed that Lesson 11 outscored Transfer lessons only in Lesson Design and Procedural Knowledge (Table 3). Alternatively, Lesson 11 outscored Baseline and Lesson 7 in all subcategories except Propositional Knowledge. We speculate that the engineering design structure of Lesson 11 allowed teachers to score significantly higher in the subcategories of Lesson Design and Procedural Knowledge, as aspects of design and procedure were made explicit within the lesson plans and are central to a quality engineering-focused lesson. Lesson 11 scores were not significantly higher than Transfer lesson scores in the subcategories of Classroom Culture and Teacher–Student Relationships (Table 3), which we attribute to the successful transfer of pedagogical skills, possibly as a result of teachers' participation in the educative curriculum-based PD. This hypothesis is further supported by the fact that teachers' Transfer lessons also scored significantly higher than teachers' Baseline lessons in the subcategories of Classroom Culture and Teacher–Student Relationships.

Performance on Lesson 11 did not significantly exceed that of any other lesson in the subcategory of Propositional Knowledge (Table 3). This suggests that teachers had a solid foundation in the content related to their selected (Baseline and Transfer) or assigned (Hemodialysis Lessons 7 and 11) lessons. That is, teachers likely consciously shared lessons that were rich in the STEM content they were comfortable teaching, which resulted in high Baseline (and Transfer) Propositional Knowledge scores. Notably, there was not much room for pedagogical improvement within this subcategory. Similarly, both INSPIRES Lessons 7 and 11 were designed to be rich in STEM content and may yield comparably high scores in Propositional Knowledge when instructed as intended.

Subcategorical and overall RTOP comparisons between Baseline lesson and Lesson 7 performance revealed no significant differences (Table 3). As part of the INSPIRES Hemodialysis curriculum, Lesson 7 is written to be rich in STEM content and also reformed in the suggested pedagogy of the STEM process. For example, teachers are encouraged to allow students to select their own independent variables, develop and justify their own predictions, design their own procedure, and share their findings with the class. Since RTOP scores did not indicate growth in pedagogical reform between the Baseline lesson and Lesson 7 enactment, we speculate that several teachers may have veered from the INSPIRES lesson plan. Forthcoming discussion of the qualitative findings explains the traits of Lesson 7 enactment that may have hindered pedagogical growth at this time point.

The present study addresses whether growth in pedagogical reform is evident in teacher-selected and teacher-written lessons (i.e., the Transfer lessons). Indeed, significant growth occurred between the Baseline and Transfer lessons, in both overall and the subcategories of Classroom Culture and Teacher–Student Relationships. These subcategories assess the degree to which teachers act as patient facilitators while creating a classroom environment that invites student communication, divergent thinking, active participation, and other qualities of student-directed learning (Piburn and Sawada 2000). We speculate that growth in teachers' pedagogical reform was influenced by their participation in the INSPIRES PD institute and the subsequent enactment of the Hemodialysis curriculum, which incorporates several pedagogical skills valued on the RTOP scale. Future discussion of qualitative findings helps identify common pedagogical traits that explain growth in the areas of Classroom Culture and Teacher–Student Relationships. Over the course of this longitudinal research study, we will make comparisons between the teacher participants and a group of teachers in a control group, which will better enable us to draw causal conclusions about the effects of the combined PD and educative curriculum on pedagogical growth.

Finally, the quantitative results indicate that comparisons between biology and technology education teachers' performance did not yield any significant differences. This finding was surprising, as we speculated that biology teachers may be stronger than technology education teachers in enactment of the science-rich Lesson 7. Likewise, we thought technology education teachers may be stronger than biology teachers in the enactment of engineering-rich Lesson 11. These assumptions may still be true, as the RTOP scale may not be the instrument that can best capture this content-specific difference. That is, the RTOP instrument measures levels of pedagogical reform in STEM fields but does not necessarily differentiate between specific STEM domains. The forthcoming exploration of qualitative trends reveals some indication that despite experience and strong content knowledge in science, biology teachers do not always teach biology lessons using reformed pedagogy and may not have followed the Lesson 7 plan as written; similarly, even with experience and a background in designing and building projects, technology education teachers do not always incorporate reformed pedagogy when teaching the engineering process.

Qualitative Findings

Qualitative analysis was explored to explain and enhance the quantitative findings of the RTOP instrument. As lessons progressed longitudinally, teachers provided more prescribed, guided parameters within Baseline lessons and Lesson 7 (Tables 4 and 5, *Guided vs. Open Strategy*) and then progressed to allowing open-ended and autonomous elements within Lesson 11 and Transfer lessons (Tables 6 and 7, *Guided vs. Open Strategy*). The nature of Lesson 11 (as written) supported the open-ended design of a hemodialysis system which likely allowed this lesson to score significantly higher than others on the RTOP scale. Although Lesson 7 was written to allow student autonomy, we found that both biology and technology education teachers often controlled the lesson by presenting vocabulary prior to the experiment, telling/assigning independent variables to student groups, providing explicit procedures, and doing mathematical computations for students. It is not surprising that even the biology teachers altered Lesson 7 in these ways, which are common practices in traditionally taught science lessons, and some level of prior pedagogical discontentment may be necessary to motivate teachers to adopt reformed methodology (e.g., Southerland et al. 2012; McNeill et al. 2017). Such reworking of the Lesson 7 plan may account, in part, for why the RTOP analysis did not reveal a significant difference between biology and technology education teacher performance (Table 2) and, more generally, why quantitative RTOP scores are relatively low for Lesson 7 enactments. As a longitudinal study with progressive PD and experience implementing the INSPIRES Hemodialysis curriculum, we predict that the reformed qualities of Lesson 7 enactment may improve in subsequent years and that increased RTOP measures would naturally follow. Research in the field of PD programs has demonstrated that increasingly difficult changes in practice (e.g., biology teachers infusing engineering practices and technology education teachers incorporating scientific rationale) require increased PD time. Further, teachers evolve their practices differently over time and therefore require flexible instructional support to continue their pedagogical and content knowledge growth (Luft and Hewson 2014). The INSPIRES PD Institute takes a learner-centered approach, where teachers are the learners and their needs guide the focal topics for continued PD sessions over the course of a three-year study. Although Transfer lessons were typically not as strongly reformed as Lesson 11 (Table 3), we have found that Transfer lessons still incorporate more aspects of autonomy than Baseline lessons, such as students guiding the procedure instead of the teacher, and more emphasis on student rationale rather than the teacher telling key ideas (Tables 4 and 7, *Guided vs. Open Strategy*). This suggests that (1) these qualitative elements may account for some of the significant growth in RTOP scores between Baseline and Transfer lessons and (2) teacher participation in the INSPIRES PD and educative curriculum enactment may influence their pedagogical growth.

The absence of pedagogical growth between Baseline lessons and Lesson 7 (Table 3) may also be attributed to how teachers probed for prior student knowledge over the course of the four documented time points. There was a tendency to utilize traditional bell work (i.e., drills to review prior concepts and completed individually; elicited student knowledge does not change the focus or sequence of the day's lesson) in both Baseline lessons and Lesson 7 (Tables 4 and 5, *Probing of Prior Knowledge*). By Lesson 11, teachers more frequently utilized reformed methods of eliciting prior knowledge (e.g., student artifacts from previous lessons) that typically progressed into a whole class discussion of scientific concepts relevant to aiding students in the next steps of their design challenge (Table 6, *Probing of Prior Knowledge*). A more widespread use of artifact sharing was observed; this pedagogical technique was explicitly modeled and encouraged during all INSPIRES PD sessions. Proper artifact sharing challenges students to make connections between the STEM concept underlying their chosen artifact and the greater design challenge of the Hemodialysis unit (Blumenfeld et al. 1991; Singer et al. 2000; Krajcik 2015). Transfer lessons often avoided traditional bell work and generally engaged students (Table 7, *Probing of Prior Knowledge*), although employed strategies were not as reformed as Lesson 11 (i.e., student presentation of prior results in lieu of artifacts).

The use of "prediction making" revealed qualitative differences in the areas of student sharing and student rationale. That is, Baseline lessons treated predictions as formalities in the scientific process while enactment of Lesson 7 posed predictions as a confirmational strategy, yet during both lessons teachers did not typically ask students to share their predictions or provide scientific rationale (Tables 4 and 5, *Making Predictions*). Alternatively, most teachers expected students to share their design ideas within groups and with the teacher during Lesson 11. Students were also expected to provide scientific or mathematic rationale for their design decisions (Table 6, *Making Predictions*). Since Lesson 11 is engineering-based, the authors treated "designs with rationale" as well-constructed predictions, as they demonstrate students' justified belief that their idea will succeed. Transfer lessons were typically more reformed in the area of students sharing their predictions with the teacher; yet, in general, the press for rationale was shallow (Table 7, *Making Predictions*). However, this gradual improvement in reformed pedagogy may help explain why the overall RTOP scores demonstrate growth from Baseline to Transfer lessons (Table 3).

When considering "connection making" within lessons, there is some level of (1) "past-to-present" and (2) "real world" connection evident at all four time points (Tables 4–7, *Making Connections*). That is, teachers commonly revisited concepts, data, etc., from previous lessons and helped students apply that prior knowledge to the current lesson. Lesson 11 continued to stand out, however, in that teachers encouraged students to make more explicit connections between multiple STEM domains (e.g., connecting science concepts to engineering design decisions) and more frequently referenced the engineering design loop and design challenge requirements (criteria and constraints). Connections to the overall engineering design challenge during Lesson 7 were typically superficial.

Qualitative findings suggest that the INSPIRES lessons were more conducive to student reflection than either the Baseline or Transfer lessons. Both Lesson 7 and Lesson 11 encouraged students to use a journal to record and reference scientific and engineering concepts. However, journals were typically used as documentation tools during the science-based Lesson 7 (Table 5, *Student Reflection*). Student reflection on how Lesson 7 could inform their approach of the engineering design challenge was limited. Notably, teachers often ran short on time during Lesson 7 and could not include all concluding elements of the lesson plan in a single 90-min period. This often played out in students not finishing their experiments, teachers stepping in to do mathematical computations for students, and teachers announcing that class-wide experimental findings would be discussed in a future class (qualitative data not shown). Although the alterations some teachers made to the INSPIRES Lesson 7 plan might influence the duration of the lesson (see discussion on *Open vs. Guided Strategy* in the previous texts), it is understandable that the absence of result sharing, interpretation, and application would confer a lower score on the RTOP scale. According to the literature, when teachers engage their students in an inquiry-based lesson, sometimes more focus is placed on completing the activity correctly than on taking the proper steps to assist students' understanding of the underlying STEM concepts (Blumenfeld et al. 1991; Singer et al. 2000). Reserving time to connect the activity to concepts during the introduction and conclusion of the lesson is an approach outlined in all lessons of the INSPIRES Hemodialysis unit. Commonly, teachers would alter the lesson plan by front-loading information (i.e., vocabulary review) before the inquiry-based lab activity of Lesson 7. Consequently, many teachers did not have time to complete the experiment and/or engage in a deep reflection at the conclusion of the period. Student reflection during Lesson 11 was enhanced as journal use became more dynamic. Engineering journals served as a forum for critical thinking in addition to documentation (Table 6, *Student Reflection*). Baseline and Transfer lessons yielded shallow student reflections centered around teacher-prompted recollection of facts at the end of the lesson (Tables 4 and 7, *Student Reflection*).

One reason why Lesson 11 may be more reform-oriented than the other lessons is because the design-based lesson may have pushed teachers from their comfort zones and encouraged them to follow the lesson plan more closely. Evidence for this speculation is presented when teachers enact specific pedagogical strategies in Lesson 11, but not Lesson 7, although such strategies are outlined in both lesson plan guides. For example, artifacts are explicitly encouraged in the guides for both Lessons 7 and 11; we observed teachers enacting student artifact-sharing more in Lesson 11 than in Lesson 7. Similarly, both lesson plan guides encourage teachers to prompt students in sketching their experimental systems. Within our qualitative subsample, we found that only technology education teachers followed this strategy during Lesson 7, while both biology and technology education teachers prompted design sketches in Lesson 11. In the latter example, technology education teachers may have followed the Lesson 7 plan more closely than the biology teachers, perhaps because the non-science teachers require more support while enacting a science-based lesson. Then, perhaps all teachers sought extra support from the Lesson 11 guide when enacting a novel, engineering design-based lesson. Therefore, while there were no quantitative significant differences identified between technology education and biology teacher RTOP scores, the qualitative analysis suggests that technology education teachers may have been following the lesson plan more closely than biology teachers during Lesson 7. Anecdotal evidence, based on conversations with multiple biology teacher participants during the INSPIRES summer PD institute, revealed that several of these teachers had previously instructed lab-based lessons on the concept of diffusion. Although the underlying concept of diffusion and some of the materials (e.g., dialysis membrane) may be similar between the INSPIRES Hemodialysis Lesson 7 and a traditional high school biology diffusion lab, the overall structure and supportive pedagogy were likely very different. Often, traditional labs are conducted as confirmational activities where information is front-loaded, rather than opportunities to exercise students' ability to think critically. Although Lesson 7 is framed as an inquiry-based lesson, its structure may have been traditionalized if science teachers felt they had enacted similar diffusion labs before and therefore reverted to the traditional strategies they used to teach a typical diffusion lab lesson. That is, if teachers believe they are enacting something familiar or do not recognize the need for, or nuance in, the reform (i.e., conducting the lab in a different manner to highlight different practices), then there may be less motivation to adjust an existing schema of how-to-teach a seemingly familiar lesson (e.g., Southerland et al. 2012; McNeill et al. 2017).

Lesson 11 was the only documented time point where teachers shared their personal experiences with students of grappling with the INSPIRES Hemodialysis unit (Table 6, *Teacher Sharing*). By conveying their personal struggles, teachers brought a humanizing component to their teaching and the lesson. Teachers and students could relate in their experience of a challenging open-ended problem. By relating to the students as they wrestled with the project, teacher–student bonds may have been established that in turn could influence students' persistence, as teacher–student relationships and teacher empathy have positive influences on student learning outcomes (e.g., Faber and Mazlish 2008; Jennings and Greenberg 2009). During the summer PD institute, many teachers voiced concerns over their students' fragility over failure and the INSPIRES unit presenting too great of a challenge for students' self-esteem. Previous research has shown that students of varied abilities are capable of success in open-ended design challenges similar to the INSPIRES Hemodialysis unit (Reeves and Ross 2010), although teachers often underestimate students' abilities to pursue and learn from these challenges (e.g., Bryan and Atwater 2002). Other research on the use of educative curricula has shown that teachers' approaches to teaching science are transformed (Pringle et al. 2017), and perhaps, the INSPIRES teachers are beginning to transform their methodology based on their experience working through the curriculum. Relating experiences of struggles and persistence to even small victories may have supported or maintained student confidence and participation for the duration of Lesson 11.

One of the questions that the present study posed was whether a shift toward reformed pedagogy would be evident between Baseline and Transfer lessons. Indeed, quantitative analyses have revealed that such a shift has begun, especially in the areas of Classroom Culture and Teacher–Student Relationships (Table 3). Qualitative analyses further explain how teachers demonstrate growth in these specific areas (Tables 4 and 7). In particular, there is an increase in elicited student ideas, student engagement, communicating (shallow) rationale with teachers, student autonomy, and (implicit) connections to data or concepts of prior lessons. The RTOP subcategories of Classroom Culture and Teacher–Student Relationships assess the degree to which teachers act as patient facilitators while creating a classroom environment that invites student communication, divergent thinking, active participation, and other qualities of student-directed learning (Piburn and Sawada 2000). Therefore, the qualitative evidence that characterizes typical Baseline and Transfer lessons supports the significant quantitative gains observed in these domains. The present study documents teacher growth after one year of participation in a three-year longitudinal study. Continued participants will experience two subsequent summer INSPIRES PD institutes, spanned by multiple monthly PD sessions. Therefore, we predict that this extended PD model will support increased growth in pedagogical reform over the final two years of the study. Substantial and difficult change in practice and content knowledge requires an increased commitment to PD-based support (Luft and Hewson 2014).

Overall pedagogical growth between Baseline and Transfer lessons may be further supported by increased incidence of argumentation. In Lesson 11 and Transfer lessons, teachers typically set higher standards of pressing students for providing STEM-based rationale. Previous research in argumentation within STEM classrooms has documented significant gains in both the frequency and quality of arguments between the first and second years of implementation (Erduran et al. 2004). Yet, in a separate study, Osborne et al. (2004) found that teachers' participation in a argumentation-focused PD program that ran 3–6 h once a month for nine months influenced growth in the quality of students' arguments, albeit not significantly. It is thought that recurrent argumentation throughout the curriculum would better support significant growth in the skill, rather than argumentation occurring primarily during nine lessons taught over the nine-month period. In the INSPIRES unit, teachers are encouraged to incorporate argumentation in multiple lessons and are further supported in developing this skill throughout three consecutive, annual, week-long summer PD institutes spanned by multiple 2-h-long monthly PD sessions. Thus, there is great potential that argumentation will grow significantly by the end of the longitudinal study. McNeill et al. (2017) supported a group of middle school science teachers in enacting an educative curriculum focused on improving argumentation; they found that while some teachers used instructional practices in line with argumentation, several others oversimplified the structured curriculum, which resulted in traditionally led lessons where students engaged in *pseudoargumentation*. Those teachers that best supported their students in developing argumentation discourse were those that (1) understood argumentation to be a cognitively enriching process, (2) actively reflected on the educative curriculum, and (3) exhibited discontent with their prior teaching methodology. Similarly, Marco-Bujosa et al. (2017) found that teachers who openly engaged in their own learning, while enacting an educative curriculum, made larger learning gains in argumentation practices than those teachers that treated the educative curriculum primarily as a resource for student activities. Therefore, INSPIRES participants may benefit from ongoing PD opportunities to actively reflect on their growth in reformed pedagogy. At this time, argumentation witnessed in the INSPIRES classrooms somewhat resembles Osborne et al.'s (2004) and McNeill et al.'s (2017) findings, as much of the teacher press and student rationale observed during Transfer lessons was present yet shallow in quality, and discourse quickly ended following students' superficial contributions. Parallel work has utilized instruments to document teachers' self-reported engineering self-efficacy and areas of concern, longitudinally over the course of the three-year INSPIRES project, which may shed light on which teachers felt discontent with their practices at different stages of the study. Finally, McNeill and Knight (2013) found that classroom argumentation was significantly enhanced following a PD program that included the following components: (1) analyzing evidence of prior classroom practice, (2) supporting teachers in infusing argumentation within lessons, (3) expecting teachers to share selected evidence of their classroom practice, and (4) encouraging teacher reflection on past practices to modify practices for the future. The INSPIRES PD institute also captures elements of these four themes as it includes the following: (1) documentation and analysis of baseline-level teacher practices (as described in the present study), (2) continued discussion and modeling of how teachers can press students for scientific and quantitative rationale for design decisions, (3) requesting that teachers prepare and share artifacts from their recent infusion of reformed pedagogical strategies, and (4) creating space for reflection and setting new goals during monthly PD sessions. Once again, the deliberate planning of the INSPIRES PD program alongside the careful structuring of the educative curriculum holds promise for substantial teacher growth and student learning.

Conclusion

Overall, we find that results addressing our first research question demonstrate that reformed pedagogy improved significantly during the first year of the study. Particularly, the instructional practices of the teachers improved significantly between enactment of the Baseline and Transfer lessons during the first year of the PD program. The findings are well aligned with previous studies when a similar PD model was utilized with middle school science teachers (Singer et al. 2011) and with high school technology education teachers (Singer et al. 2016). Both prior studies used a similar repeated measures design to analyze RTOP scores. Results from the present study were conducted with a much larger population of teachers and also demonstrated significant differences on more RTOP subcategories than prior studies. Unlike the present study, Singer et al. (2016) found gains in Propositional Knowledge. Video coders noted that while teachers enacted the INSPIRES curriculum, the teachers often failed to connect the design challenge (building a hemodialysis machine) to the science concepts (e.g., diffusion). Student ideas were often solicited then discarded for the teachers' preconceived ideas of how the lesson should proceed. By Lesson 11, teachers began releasing control of the lesson direction to students and allowed them to design and build their own machines. Even with stronger emphasis on student ideas, connections to the underlying STEM practices were inconsistent. After-school PD meetings used a lesson-study model and fostered discussions about how to connect the science and engineering more strongly to Lesson 11 and how to lead other lessons more similarly to Lesson 11.

Qualitative analysis demonstrated that Transfer lessons exhibited more reformed qualities (i.e., student autonomy, connections to prior knowledge, open-ended design-based activities) than Baseline lessons. Multiple themes emerged that were used to characterize each lesson: Baseline, Lesson 7, Lesson 11, Transfer (Tables 4–7).

Regarding our second research question, we do not see a significant difference between biology and technology education teachers' pedagogical growth at this time. We recognize that this finding may change as this research project continues to unfold. The following two years of this longitudinal study are expected to yield further reform in pedagogical skills and the integration of engineering practices into STEM classrooms. Close observation of this pedagogical evolution has the potential to reveal differences between the biology and technology education teacher populations that may surface at later times. To date, these findings provide insights for rethinking the structure of professional development, particularly in the integrated use of an educative curriculum aligned with intended professional development goals.

Recommendations

Results from the present study will be compared against RTOP data and qualitative trends measured from teachers in a control group. The control group comprises biology and technology education teachers in the same district who did not participate in the INSPIRES PD or implement the INSPIRES curriculum. Additionally, while the RTOP rubric facilitated the present study of student-centered pedagogical change in STEM classroom environments, other observational tools exist that more specifically address changes in classroom engineering practices and principles. The teacher lessons evaluated here via the RTOP were simultaneously coded using a research instrument sensitive to explicit engineering lesson qualities. Next steps in research include the analysis and dissemination of forthcoming findings pertaining to engineering-specific changes and how they may align to the broader RTOP results. In general, we recommend that educative curricula be used as a vector for integrating elements of educational reform to address NGSS challenges, especially in engineering education. Professional development that supports teaches in implementing a strongly written engineering educative curriculum can allow the transfer of design-based pedagogy into teacher-developed curricula.

Acknowledgements We wish to thank the teachers, administrators, and program staff of our partner school district for their time and efforts in this study. The following UMBC students assisted in the collection of classroom recordings: Abby Singer, Ahmed Al-Salihi, Goureesh Paranjpe, Garrett Bockmiller, Marcus Foster, and Ekaterina DiBenedetto.

Funding Information This project was funded by a Discovery Research K–12 National Science Foundation grant (DRL 1418183).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Amolins, M. W., Ezrailson, C. M., Pearce, D. A., Elliott, A. J., & Vitiello, P. F. (2015). Evaluating the effectiveness of a laboratory-based professional development program for science educators. *Advances in Physiology Education, 39*(4), 341–351.

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

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

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3–4), 369–398.

Bryan, L. A., & Atwater, M. M. (2002). Teacher beliefs and cultural models: a challenge for science teacher preparation programs. *Science Education, 86*(6), 821–839.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: building networked improvement communities in education. In *Frontiers in sociology of education* (pp. 127–162). Dordrecht: Springer.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluation normed and standardized assessment instruments in psychology. *Psychology Assessment, 6*(4), 284–290.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: complementary research strategies* (pp. 54–82). Thousand Oaks: Sage Publications.

Cuban, L. (2013). *Inside the black box of classroom practice: change without reform in American education*. Cambridge: Harvard Education Press.

Darling-Hammond, L., & Mclaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*, 597–604.

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

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

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.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*(6), 915–933.

Faber, A., & Mazlish, E. (2008). *How to talk so kids can learn*. New York: Simon and Schuster.

Fishman, E. J., Borko, H., Osborne, J., Gomez, F., Rafanelli, S., Reigh, E., Tseng, A., Million, S., & Berson, E. (2017). A practice-based professional development program to support scientific argumentation from evidence in the elementary classroom. *Journal of Science Teacher Education, 28*(3), 222–249.

Knaggs, C., & Schneider, R. (2012). Thinking like a scientist: using vee-maps to understand process and concepts in science. *Research in Science Education, 42*(4), 609–632.

Krajcik, J. S., & Blumenfeld, P. C. (2006). *Project-based learning* (pp. 317–334).

Krajcik, J. (2015). Project-based science: engaging students in three-dimensional learning. *The Science Teacher, 82*(1), 25.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*(1), 491–525.

Lotter, C., Rushton, G., & Singer, J. E. (2013). Teacher enactment patterns: how can we help move all teachers to reform-based inquiry practice through professional development? *Journal of Science Teacher Education, 24*, 1263–1291.

Loucks-Horsley, S., Stiles, E., Mundry, S., Love, N., & Hewson, P. (2010). *Designing professional development for teachers of science and mathematics* (3rd ed.). Thousand Oaks: Corwin.

Luft, J. A., & Hewson, P. W. (2014). Research on teacher professional development programs in science. *Handbook of Research on Science Education, 2*, 889–909.

MacIsaac, D., Sawada, D., & Falconer, K. (2001). *Using the Reformed Teaching Observation Protocol (RTOP) as a catalyst for self-reflective change in secondary science teaching*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

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

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in the middle grades: assessment of learning in urban systemic reform. *Journal of Research in Science Teaching, 41*(10), 1063–1080.

McAleer, S. D. (2008). Professional growth through mentoring: a study of experienced mathematics teachers participating in a content-based online mentoring and induction program. Dissertation Abstracts International-A, 69 (08).

McNeill, K. L., González-Howard, M., Katsh-Singer, R., & Loper, S. (2017). Moving beyond pseudoargumentation: teachers' enactments of an educative science curriculum focused on argumentation. *Science Education, 101*(3), 426–457.

McNeill, K. L., & Knight, A. M. (2013). Teachers' pedagogical content knowledge of scientific argumentation: the impact of professional development on K–12 teachers. *Science Education, 97*(6), 936–972.

National Research Council. (2012). *A framework for K-12 science education: practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press Retrieved from http://www.nap.edu/catalog.php?record_id=13165.

Next Generation Science Standards. (2013). *Next Generation Science Standards. [Website]*. Washington, DC: National Research Council, National Science Teachers Association, & American Association for the Advancement of Science Retrieved from http://www.nextgenscience.org/.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020.

Piburn, M., & Sawada, D. (2000). *Reformed teaching observation protocol (RTOP): reference manual (ACEPT technical report IN00-3)*. Tempe: Arizona State University, Arizona Collaborative for Excellence in the Preparation of Teachers.

Pringle, R. M., Mesa, J., & Hayes, L. (2017). Professional development for middle school science teachers: does an educative curriculum make a difference? *Journal of Science Teacher Education, 28*(1), 57–72.

Reeves, R., & Ross, J. (2010). AC 2010–1952: a novel approach to professional development. *American Society for Engineering Education, Conference Proceedings*, 15, 1.

Reiser, B. J. (2014). *Designing coherent storylines aligned with NGSS for the K-12 classroom*. Paper presented at the Professional Development Institute of the National Science Education Leadership Association, Boston, MA. Retrieved from https://www.academia.edu/6884962/Designing_Coherent_Storylines_Aligned_with_NGSS_for_the_K-12_Classroom. Accessed 22 June 2017.

Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *The Elementary School Journal, 100*(4), 331–350.

Richmond, G., Parker, J. M., & Kaldaras, L. (2016). Supporting reform-oriented secondary science teaching through the use of a framework to analyze construction of scientific explanations. *Journal of Science Teacher Education, 27*(5), 477–493.

Roseman, J. E., Fortus, D., Krajcik, J., & Reiser, B. J. (2015). *Curriculum materials for Next Generation Science Standards: what the science education research community can do*. In NARST Annual International Conference, Chicago, IL.

Roseman, J. E., Herrmann-Abell, C. F., & Koppal, M. (2017). Designing for the Next Generation Science Standards: educative curriculum materials and measures of teacher knowledge. *Journal of Science Teacher Education, 28*(1), 111–141.

Ross, J., Bayles, T., & Singer, J. (2015). The inspires curriculum. In C. Sneider (Ed.), *The go-to guide for engineering curricula, grades 9–12* (pp. 19–30). London: SAGE Publications Ltd. https://doi.org/10.4135/9781483388373.n3.

Rushton, G., Lotter, C., & Singer, J. (2011). Chemistry teachers' emerging expertise in inquiry teaching: the effect of an authentic professional development model on beliefs and practice. *Journal of Science Teacher Education, 22*, 23–52.

Saderholm, J., Ronau, R. N., Rakes, C. R., Bush, S. B., & Mohr-Schroeder, M. (2016). The critical role of a well-articulated conceptual framework to guide professional development: an evaluation of a state-wide two-week program for mathematics and science teachers. *Professional Development in Education*. https://doi.org/10.1080/19415257.2016.1251485.

Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: the range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching, 42*(3), 283–312.

Singer, J., Marx, R. W., Krajcik, J., & Clay Chambers, J. (2000). Constructing extended inquiry projects: curriculum materials for science education reform. *Educational Psychologist, 35*(3), 165–178.

Singer, J., Lotter, C., Feller, R., & Gates. (2011). Exploring a model of situated professional development: impact on classroom practice. *Journal of Science Teacher Education, 22*(3), 203–227.

Singer, J. E., Ross, J. M., & Jackson-Lee, Y. (2016). Professional Development for the Integration of Engineering in High School STEM Classrooms. *Journal of Pre-College Engineering Education Research, 6*(1), 30–44.

Southerland, S. A., Nadelson, L., Sowell, S., Saka, Y., Kahveci, M., & Granger, E. M. (2012). Measuring one aspect of teachers' affective states: development of the science teachers' pedagogical discontentment scale. *School Science and Mathematics, 112*(8), 483–494.

Spradley, J. P. (1980). *Participant observation*. Orlando: Harcourt Brace Jovanovich College Publishers.

Sztajn, P. (2011). Research commentary: standards for reporting mathematics professional development in research studies. *Journal for Research in Mathematics Education, 42*(3), 220–236.

Willis, M. B. (2018). Examination of novice science teachers' use of project-based instructional strategies (Doctoral dissertation, University of Massachusetts Lowell).