ISSN: 2148-9955

International Journal of Research in Education and Science (IJRES)

www.ijres.net

A High Quality Educative Curriculum in Engineering Fosters Pedagogical Growth

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

[1]Department of Education, University of Maryland Baltimore County, USA
[2]College of Engineering, Virginia Tech, USA

To cite this article:

Williams, T., Krikorian, J., Singer, J., Rakes, C., & Ross, J. (2019). A high quality educative curriculum in engineering fosters pedagogical growth. *International Journal of Research in Education and Science (IJRES), 5*(2), 657-680.

International Journal of Research in Education and Science

Volume 5, Issue 2, Summer 2019 ISSN: 2148-9955

A High Quality Educative Curriculum in Engineering Fosters Pedagogical Growth

Tory Williams, Jacqueline Krikorian, Jonathan Singer, Christopher Rakes, Julia Ross

Article Info

Article History

Received:
12 April 2019

Accepted:
26 July 2019

Keywords

Educative curriculum
Engineering education
Mixed methods
Pedagogical reform
Professional development

Abstract

The Next Generation Science Standards call for the integration of engineering into mainstream science education, yet challenges arise for teachers unfamiliar with engineering-based pedagogy or concepts. The INSPIRES Hemodialysis educative curriculum may address such challenges by explicitly integrating all areas of STEM in an authentic learning experience with pedagogical support for teachers. Preparation and implementation of the INSPIRES curriculum is paired with professional development (PD) guided by the PrimeD framework. The present three-year study explored the role of the INSPIRES curriculum and PD in strengthening teacher pedagogical skills while implementing engineering ideas and practices in high school biology and technology education classrooms. Teachers' classroom practices were measured with the RTOP and qualifying themes emerged through further analyses across multiple time points. Initially, PD focused on the INSPIRES curriculum but later supported teachers in individual areas of concern. Growth in design-based pedagogy was evident in both engineering-rich INSPIRES lessons and subsequent teacher-developed lessons which highlighted the transfer of new skills. While skill transfer occurred within the teacher population discussed here, sustained pedagogical reform may require continued PD support over time. Overall, educative curricula may provide a vector for integrating elements of educational reform to address NGSS challenges, especially in engineering education.

Introduction

The current landscape of secondary STEM pedagogy is amidst profound changes in response to the adoption of *Next Generation Science Standards* (NGSS). The NGSS framework embraces the integration of Scientific and Engineering Practices, Crosscutting Concepts, and Disciplinary Core Ideas ("3D Learning"). That is, STEM pedagogy ought to introduce "key organizing concepts" (Disciplinary Core Ideas) along with related, practical skills (Practices) and application of patterns across multiple disciplines (Crosscutting Concepts; NGSS, 2013). An emphasis on engineering education is a natural fit for 3D Learning, as engineering content and practices (as highlighted in Carr, Bennett IV, & Strobel, 2012) call for the application of STEM-based skills and concepts toward solving real-world problems. The inclusion of engineering ideas and practices into NGSS clearly initiates a new shift toward the integration of engineering into mainstream science education (National Research Council, 2012; Next Generation Science Standards, 2013). Yet, this shift presents significant challenges to teachers unfamiliar with engineering-based pedagogy and engineering or science concepts.

Over the past twelve years, the integration of engineering content into K-12 classrooms has increased. At the elementary level, the *Engineering is Elementary* curriculum has achieved broad national distribution, impacting over 4 million students to date (Museum of Science, 2012). In grades 9-12, the introduction of engineering concepts has occurred primarily through technology education classrooms. While technology education once encompassed primarily vocational skills, many states are now implementing strategies to broadly expose their students to technology through pre-engineering experiences. Several curricula have achieved national-level distribution including Project Lead the Way and Engineering by Design. Project Lead the Way provides a sequence of pre-engineering courses targeting students considering engineering careers (Project Lead the Way, 2016). Engineering by Design provides whole-year course curricula for technology education targeting technological literacy for all

(International Technology & Engineering Educators Association, 2016). Despite these significant advances, engineering has not yet become integrated as a mainstream content area nationally (National Research Council, 2009). Additionally, after studying the outcomes of Project Lead the Way and other existing engineering curricula, Kelley, Brenner and Pieper (2010) found that students used a maximum of 3% of their time applying quantitative reasoning to guide their design decisions. Thus, there exists a need for curriculum writers to take increased consideration into developing learning opportunities that integrate quantitative thinking into science inquiry activities while enacting the engineering design process (Kelley et al., 2010).

The INcreasing Student Participation, Interest, and Recruitment in Engineering and Science (INSPIRES) curriculum was written to explicitly integrate all areas of STEM for an authentic learning experience in high school classrooms Singer, Ross, & Jackson-Lee, 2016). INSPIRES considers the success of diverse learners, and lessons are purposefully designed to be inclusive of all students by engaging in engineering principles and practices. In addition, the INSPIRES curriculum has been carefully constructed to be educative for teachers (as recommended by Ball & Cohen, 1996; Schneider, Krajcik, & Blumenfeld, 2005; Knaggs & Schneider, 2012). The integrated use of educative curriculum materials provides support for teachers by including features that encourage reflection and promote connections among specific content, pedagogy and pedagogical content knowledge. INSPIRES units combine real-world engineering design challenges with inquiry-based learning strategies to engage students as they learn to apply scientific, quantitative, and technological thinking to solve a problem. In turn, INSPIRES units have the potential to increase students' technological literacy while supporting the development of skills that foster success in STEM disciplines.

The INSPIRES curriculum was designed to be relevant to multiple STEM classrooms (i.e., science or technology education), low in cost, and short in duration (approximately three weeks in length (Ross, Bayles, & Singer, 2015; Singer et al., 2016). Elements of STEM Practices, Crosscutting Concepts, and Disciplinary Core Ideas are infused throughout multiple lessons. Specifically, the INSPIRES curriculum addresses all four NGSS Engineering Design performance expectations (HS-ETS1) and all eight Scientific and Engineering Practices (NGSS, 2013). These features allow INSPIRES to offer unique and authentic learning opportunities for teachers and students of grades 9-12 STEM classrooms (Ross et al., 2015; Singer et al., 2016).

The present study explored the role of the INSPIRES educative curriculum, and accompanying professional development (PD), in strengthening teacher pedagogical skills while implementing engineering ideas and practices in high school biology and technology education classrooms. Initial results indicated that pedagogical growth was evident after the first year of intervention (Williams, Singer, Krikorian, Rakes, & Ross, 2019). The full effect of the complete, three-year, intervention is examined in the present longitudinal study. The research questions were:

1) To what extent 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 PD accompanying the INSPIRES curriculum was guided by the Professional Development: Research, Implementation, and Evaluation (PrimeD) framework (Rakes, Bush, Mohr-Schroeder, Ronau, & Saderholm, 2017; Saderholm, Ronau, Rakes, Bush, & Mohr-Schroeder, 2017). PrimeD organizes PD into four interactive, cyclic phases: design, implementation, evaluation, and research. The design phase emphasizes the inclusion of all stakeholder groups to develop a challenge space, an explicit description of the needs, vision, goals, targets, and strategies of the PD. The implementation phase is divided into whole group activities and classroom implementation. The PrimeD elements of effective PD summarize research-based characteristics of whole group activities that have been found to improve classroom practice and student outcomes.

PrimeD envisions classroom trials as a critical intentional component of the PD, with results being brought back to subsequent whole group meetings to refine the focus, understanding, and strategies. PrimeD structures the classroom trial expectation within the PD through the Plan-Do-Study-Act (PDSA) cycle (Bryk, Gomez, & Grunow, 2011), which is grounded in improvement science (e.g., Martin & Gobstein, 2015). At the next level of organization, PrimeD structures the cycles of whole group engagement and classroom implementation through networked improvement communities (NICs; Martin & Gobstein, 2015). Ensuring that whole group engagement and classroom implementation are seen as essential and tightly connected aspects of a PD program is a key notion of PrimeD.

Specifically, INSPIRES teachers implemented activities in their classrooms that addressed one or more of the challenges discussed during a group meeting. Teachers were grouped into several smaller communities (NICs) based on geography (same or nearby school) and content area (biology or technology). Teachers then presented lesson artifacts at subsequent whole and small group meetings to drive and enrich discussion. The final research phase of PrimeD addressed the goals, design, data, threats to validity and reliability of research.

Additionally, the ongoing results of this longitudinal study continually informed the development and adjustment of the challenge space. PD facilitators worked closely with district partners and teacher participants. In addition, an external evaluator and Advisory Board served to provide critical feedback on the direction of the evolving study. The external evaluator administered multiple surveys to PD facilitators, district personnel, and participating teachers to capture various teacher attitudes, benefits and limitations of the study, and for feedback on the working relationships within the study. The Advisory Board met with the PD facilitators and district personnel on two occasions to provide guidance on the study's adherence to goals approved by the funding agency.

Characteristics of the PD

PD programs can facilitate change in classroom pedagogies that will persist as engineering concepts and practices become widely integrated into science, technology, engineering, and mathematics (STEM) classrooms (Blumenfeld, Fishman, Krajcik, Marx, & Soloway, 2000). The PrimeD Elements of Effective PD were used to frame the short- and long-term planning of PD sessions (PrimeD whole group engagement). The first element, connection to challenge space, recognizes the importance of ensuring that all group activities be guided by the design of the PD and maintaining a common vision among stakeholders, as recommended by Bryk et al. (2011). The other PrimeD elements of effective PD combine recommendations from McAleer (2008), Desimone (2009), Loucks-Horsley, Stiles, Mundry, Love, and Hewson (2010), Sztajn (2011), Putnam and Borko (2000), Borko (2004), Greeno, Collins, & Resnick (1996), Lave & Wenger (1991), Penuel, Fishman, Yamaguchi, and Gallagher (2007), and Timperley (2011). These texts represent a broad perspective on PD effectiveness, yet provide a consistent message about the nature of effective PD. The ten elements used to guide the PD toward effective practice were:

1) Connection to challenge space
2) Analysis of classroom practice
3) Analysis of student work and learning
4) Focus on specific academic content
5) Connection to school initiatives
6) Collective collaboration and respectful interactions
7) Active participation and participant leadership
8) Critical, intensive peer- and self-reflection
9) Ongoing engagement sustained over time
10) Continuous monitoring and evaluation

However, even with effective PD programs, teachers may struggle to successfully integrate engineering design- and inquiry-based practices (Schneider et al., 2005). The PrimeD feedback cycles between classroom implementation, whole group engagement, and PD design, implementation, and evaluation were critical for updating needs assessments and to adjust implementation strategies as needed.

Prior research has demonstrated that increasingly difficult changes in practice, such as infusing engineering-based pedagogy and content knowledge, requires increased PD time (e.g., Luft & Hewson, 2014). To address the PD element of ongoing engagement and sustainment over time, the amount and quality of the engagement was considered. Mehalik, Doppelt, and Schuun (2008) demonstrated that a PD program with 4-hour sessions provided before, during, and following implementation of a systems design-based science unit, results in significant gains in student content knowledge, engagement, and retention; with the largest gains observed in low-achieving African American students. Black, Harrison, Lee, Marshall, and Wiliam (2004) recommended three years as an optimal length of time to support changes in teaching; the INSPIRES PD program engaged with teachers for three years.

First year PD efforts focused on developing teachers' use of reform pedagogies (Piburn & Sawada, 2000) through the incorporation of engineering design principles. Reform pedagogies include strategies that invite collaboration, context, relationship-building, active learning, and treat the instructor as a facilitator of an ongoing learning process. The second year of PD focused on deepening teacher understanding of engineering design principles and how they support reform pedagogy. The third year of PD focused on enhancing teacher ability to synthesize their learning from the first two years and developing new engineering design challenges.

Teacher collaboration is well-accepted as being important to the general success of PD programs (as in collective collaboration and active participation/participant leadership elements of effective PD). The current gap in empirical research on PD lies in understanding *how* teachers comprehend new concepts as they participate in a PD program, and which strategies best support both teacher and student learning. Four themes that should drive the organization of PD programs include: 1) providing continued and flexible support as teachers evolve their practices; 2) frequent opportunities for teacher collaboration; 3) a coherent program; and 4) an emphasis on rich content knowledge (Luft & Hewson, 2014). One strategy to address these four requirements is the use of an educative curriculum (Ball & Cohen, 1996; Schneider et al., 2005) throughout the PD program.

Educative Curriculum

An educative curriculum is one that supports both teacher and student learning, and is a valuable tool for integrating a new content domain, such as engineering, into school districts. The pairing of educative curriculum materials with aligned PD improves pedagogical growth, as evidenced by results from small-scale studies (e.g., Rushton, Lotter, & Singer, 2011; Singer, Lotter, Feller, & Gates, 2011; Lotter, Rushton, & Singer, 2013), as well as in the prior findings of the current project (Williams et al., 2019). An educative curriculum provides a scaffold for assimilating content and reformed pedagogies during PD, as teachers first approach the curriculum from the perspective of students. In some iterations, the educative curriculum is a model that is used only in the PD meetings; in others, teachers integrate the educative curriculum with their students. When used within PD, an educative curriculum also facilitates explicit discussion of highlighted pedagogical practices and how those practices may be transferred to other lessons (Remillard, 2000). Educative curricula used in PD offer the opportunity for teachers to practice strategies for presenting abstract concepts through concrete and semi-concrete examples. Such practice has been linked to growth in content knowledge and pedagogy of high school STEM teachers (Singer et al., 2011; Lotter et al., 2013). Teachers that have implemented an educative curriculum demonstrated increased student support and expectations in practices aligned with NGSS (e.g., promoting evidence-base claims, planning next steps, etc.; Arias, Davis, Marino, Kademian, & Palincsar, 2016).

Also, teacher self-efficacy and pedagogy in science can shift in response to training and enactment of an educative curriculum (Pringle, Mesa, & Hayes, 2017). Therefore, educative curricula provide a pathway to supporting teachers in adopting NGSS (Roseman, Herrmann-Abell, & Koppal, 2017). In a manner similar to how increased time spent in PD programs yields greater shifts in teacher pedagogical or content knowledge, increased experience working with educative curricula allows teachers greater success in approaching complex problems and recognizing patterns as they arise in new contexts within the classroom environment (Noh & Webb, 2015). While educative curricula are already viewed as a rich source of engaging student lessons, more work is needed to highlight the value of these curricula as a tool for teacher pedagogical growth (Marco-Bujosa, McNeill, González-Howard, Loper, 2017).

The INSPIRES educative curriculum is a resource for integrating engineering design principles and practices into high school STEM classes. The curriculum consists of five modules: engineering in flight, engineering in the environment, engineering energy solutions, engineering in health care: heart lung, and engineering in health care: hemodialysis (Ross et al., 2015). The present study used the engineering in health care: hemodialysis to integrate engineering design principles into high school biology and technology education classes. In this module, students learn about diffusion of waste across membranes and factors that influence the diffusion rate. A project-based approach (Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palincsar, 1991), guides the student-driven design of a hemodialysis apparatus, as critical science content and engineering practices are infused along the way. The final products of the module are a variety of systems that student teams have designed, built, tested, and revised to function as a hemodialysis machine.

INSPIRES-Specific Educative Curriculum and PD Program

The combined INSPIRES educative curriculum and accompanying teacher PD provides a vector for supporting the successful adoption of NGSS practices and an appreciation for the role of applied science, math, and technology in authentic engineering problems. The PD program in the current study consisted of one-week long institutes held during three consecutive summers; and a series of two-hour after-school PD sessions, offered throughout the academic years spanning the project. The Year 1 Summer Institute (SI) focused on four key components: (1) the project's educative curriculum materials; (2) STEM practices; (3) pedagogical practices; and (4) reflective critiques. The purpose of the project's educative curriculum materials was to provide coherence among the STEM practices, pedagogical practices and reflective techniques. During the STEM practices segment of PD, the facilitators used specific activities from the materials to illustrate key ideas for deeper discussion. The project's curriculum module, which focused on designing and building a hemodialysis system, was the primary material used for this purpose within this study. The STEM practices component aimed to build content knowledge and skills, including the engineering design process, as teacher teams participated as students when engaging the design challenge materials. A design-based, phenomena-first methodology (Ball & Cohen, 1996) was emphasized through the curriculum's activities and learning objectives. The pedagogical practice component aimed to increase pedagogical content knowledge, as pedagogical strategies were explicitly modeled within the context of STEM content and practice throughout the PDs. Examples of pedagogical strategies included design-based learning (e.g., use of an engineering design loop), collaboration (e.g., Think-Pair-Share), context (e.g., driving questions, KWL charts), technology integration (e.g., simulations), and assessment for learning (e.g., probing questions). The reflective critiques component focused on pedagogical practice. The PD facilitators guided teacher-led discussions of how INSPIRES lesson structure and embedded strategies enhance engagement and content learning for students.

The Year 2 SI expanded on the aforementioned key components and also emphasized support in the transfer of pedagogical practices. The after-school, academic year PD sessions following SI Year 2 drew attention to how teachers could enhance their own lessons with strategies previously employed through the INSPIRES module. Teachers volunteered their lesson prompts as subjects of discussion during small-group PD sessions. Often, the resulting teacher-driven discussion would yield student-centered adaptations to the lesson structure. During SI Year 3, teachers focused primarily on developing their own lessons, aligning to both NGSS and new district science curricula, and the transfer of reformed pedagogical practices was emphasized. Teachers were encouraged to work in teams and to employ their newly developed or adapted lessons during the subsequent school year.

Methods

Participants

This study was conducted in collaboration with a large mid-Atlantic public school district in the United States. Students across the school district live in suburban, rural, and urban neighborhoods, and come from a cross-section of economically diverse families. Overall, 54.8% of the district's students represent minorities, 48.9% are female, and 44.8% are eligible for free/reduced price meals. Participants were recruited from 15 schools representative of the district's student diversity, and included both traditional and alternative environments for biology and technology education courses. Twenty-seven and 21 teachers participated in the treatment and control groups, respectively, during the first year of this longitudinal study (Williams et al., 2019). For the full 3-year study, 17 biology and technology education teachers (N=6 & 11 respectively) remained in the treatment group, and nine biology and technology education teachers (N = 5 & 4 respectively) remained in the control group. Attrition was therefore a threat to internal validity (as described in Shadish, Cook, & Campbell, 2002), so causes of attrition were examined throughout the study to determine possible means of prevention. The project external evaluator attempted to interview 12 teachers that had left the study early, however, the number of teachers that completed this exit interview (N=2) was considered too small to draw evaluative conclusions. Whether the causes of attrition were similar or different between the treatment and control groups is unresolved, and the potential threat to internal validity is unknown. Participants included both males (N=16) and females (N=12) who reported their race/ethnicity as Black (23%) or White (77%), and whose classroom teaching experience ranged from 2-28 years.

Teachers participating in the treatment group received an intervention in the form of INSPIRES content- and pedagogy-focused PD during summers and inter-spanning academic years as described above, as well as all materials necessary to implement the INSPIRES Hemodialysis module in one or more classrooms during the first two years of the study. Teachers that comprised the control group did not receive an intervention through the INSPIRES program. Instead, the control group provided a standard of corresponding longitudinal growth influenced by district-level PD, ongoing classroom experience, or other confounding factors that may shift how teachers approach implementation of NGSS.

The participating district's decision to reformat high school science curricula during Year 3 of this study provided an external influence on the direction of the research. The substantial change in curricula may have prompted some of the observed participant attrition (e.g., the teacher was no longer teaching biology or life sciences). Yet, a novel opportunity also arose as PD facilitators supported continuing participants in transferring the pedagogical skills acquired through INSPIRES to prepare adapted lessons for the district's new curricula. Specifically, the PD program and research agenda exercised flexibility to meet the needs of teachers. The Year 3 SI, and the academic year PD sessions leading up to that point, focused on enhancing teachers' lessons or helping teachers write new lessons, that incorporate engineering design content, practices, and the correlating supportive pedagogies.

Data Sources

The data were obtained from 90-minute (single lesson) classroom recordings at nine time points (Figure 1; * signifies those time points where data were collected from the control group, in addition to the treatment group):

1) The spring prior to Summer Institute (SI) 1, to determine baseline levels of teachers' abilities to include NGSS Engineering Design Standards (HS-ETS1) in a lesson ("Baseline");
2) After SI 1, during enactment of a science-based lesson of the INSPIRES curriculum ("SL1");
3) After SI 1, during enactment of an engineering-based lesson of the INSPIRES curriculum ("EL1");
4) After enactment of the project's curriculum module, to determine the level of transferred NGSS Engineering Design-based skills (HS-ETS1) into a teacher-developed lesson ("T1")*;
5) After SI Year 2, during enactment of a science-based lesson of the INSPIRES curriculum ("SL2");
6) After SI 2, during enactment of an engineering-based lesson of the INSPIRES curriculum ("EL2").
7) After enactment of the project's module in Year 2, to determine the level of transfer ("T2")*;
8) During 1st-quarter implementation of teacher-developed lessons, to measure transfer ("T3a")*;
9) During 2nd-quarter implementation of teacher-developed lessons, to measure transfer ("T3b")

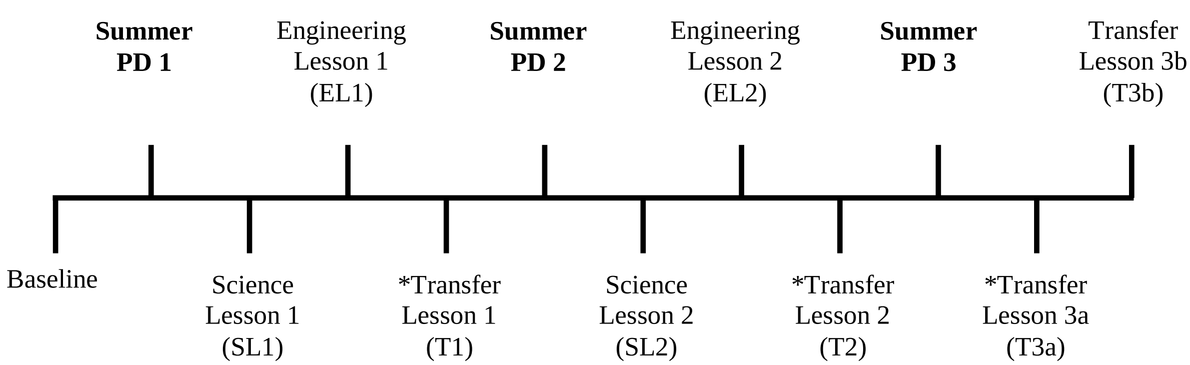

Figure 1. Timeline of study's data collection points and PDs

The treatment group participated in SI and academic year PD sessions and was sampled for data at all nine time points listed above. The control group did not participate in the project's SI or PD sessions and were only sampled for data during transfer lesson time points 4, 7, and 8, above (indicated by *). During the course of this longitudinal study, the partnering school district underwent a change in policy that required a shift from using video-recorded lessons to audio-recorded lessons in external research. Therefore, data from time points 1-3 were captured via video-recording, while data from time points 4-9 were captured via audio-recording. When notified of the policy change, the research team quickly purchased and trained in the use of professional grade audio equipment. Classroom recorders were also trained in detailed note-taking to account for nonverbal information that was previously documented by video. Further, the research team established the effectiveness of audio-recordings for data collection by finding strong similarities in lesson scoring between the classroom recorder and three scorers who could only rely on audio recordings and written notes. Inter-rater reliability was not reduced by the change from video- to audio-recording.

Classroom recordings were rated by four coders using a modified form of the Reformed Teaching Observation Protocol (RTOP; Sawada, Piburn, Judson, Turley, Falconer, et al., 2002). The use of RTOP as both a quantitative and qualitative tool is well established in STEM educational research (e.g., MacIsaac, Sawada, and Falconer 2001; Enderle, Dentzau, Roseler, Southerland, Granger, Hughes, and Saka 2014; Amolins, Ezrailson, Pearce, Elliott, and Vitiello 2015). For the present study, performance level descriptors were developed for each numeric score level (0-4) in the 25-item RTOP rubric to reduce the subjectivity of ratings. The descriptors were examined by an expert panel and then field tested. The descriptors were revised based on feedback from the expert panel and clarifications that arose from the field-testing process. At least 20% of the data were coded by multiple scorers at each data time point. Intra-class correlations were computed for each round of coding to measure and assess inter-rater reliability. Interclass correlation coefficients (K) that ranked in the range of 0.75-1.00 were considered excellent and ranks between 0.60-0.74 were considered good (Cicchetti, 1994). The mean coefficient for recordings scored by all coders was $K = 0.76$ ($\pm$ 0.11 SD). For all multi-coded recordings, discrepancies in item scores between coders were deliberated upon until consensus was reached.

The RTOP items are subdivided into five categories that each contains five rubric items: lesson design, propositional knowledge, procedural knowledge, classroom culture, and teacher-student relationships (Piburn & Sawada, 2000). Each lesson recording received a single score for each subcategory by summing the scores for its five items. An overall score was computed for all 25 items. Differences in overall and subcategories across the nine lessons were analyzed with a repeated measures analysis of variance (rmANOVA) with one fixed factor of biology vs. tech. ed. teachers. Further, rmANOVA allowed for comparisons between baseline and transfer lessons of teachers in the treatment vs. control groups.

Qualitative data was obtained from a subsample of six teacher participants; three biology and three technology education teachers. This subsample of teachers was selected based on Baseline RTOP scores that were in the mean range (+/- 1 standard error) for at least two RTOP subcategories. By selecting teacher cases whose assigned RTOP scores were around the means representative to all participating teachers, the researchers aimed to characterize the common trends in teaching practices at the different time points of the study. Further, each teacher in the qualitative subsample 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 & Atkinson, 1996).

Data Analyses

Quantitative statistical analyses involved averaging the scores for the five items per RTOP subcategory, for each teacher video or audio recording. Further, an overall average score was computed for all 25 RTOP items, per teacher recording. Differences in overall and subcategory averages across the nine lessons (time points) were analyzed with a repeated measures analysis of variance (rmANOVA) with one fixed factor to compare biology and technology education content domains. A separate rmANOVA allowed comparisons between the longitudinal performance of teachers in the treatment group and the control group on baseline and transfer lessons only. In the cases where Mauchly's Test of Sphericity demonstrated that the conditions of sphericity had not been met, the Greenhouse-Geisser estimates were used when determining statistical significance and reporting values from the rmANOVA. Pairwise comparisons between specific lesson time points were evaluated using Tukey's HSD. All statistics were performed using Microsoft Excel 2010 and IBM SPSS Statistics 21. A researcher who scored lesson videos on the RTOP scale also critically examined the RTOP scoring notes and summaries from all focal lessons enacted by the teachers in the qualitative subsample. Pedagogical traits that were generally representative of each lesson were identified and led to the development of a typical qualitative description for each lesson time point in the study. A similar method was used by Williams et al. (2019) to account for emerging qualitative themes.

Results and Discussion

Quantitative Results

Overall scores were computed for all 25 RTOP items of treatment teachers' baseline and transfer lessons, and control teachers' lessons at all three respective time points. With both teacher-groups scoring an average of approximately half the possible points on the RTOP, the initial level of pedagogical reform was similarly low in treatment (baseline lesson) and control (Year 1 lesson) teachers. Longitudinally, treatment teachers' average performance on transfer lessons in years 1, 2 and 3a are each significantly more reformed than their average baseline performance. Therefore, teachers in the treatment group showed significant growth from their own starting point, as well as from the starting point of teachers in the control group (Figure 2).

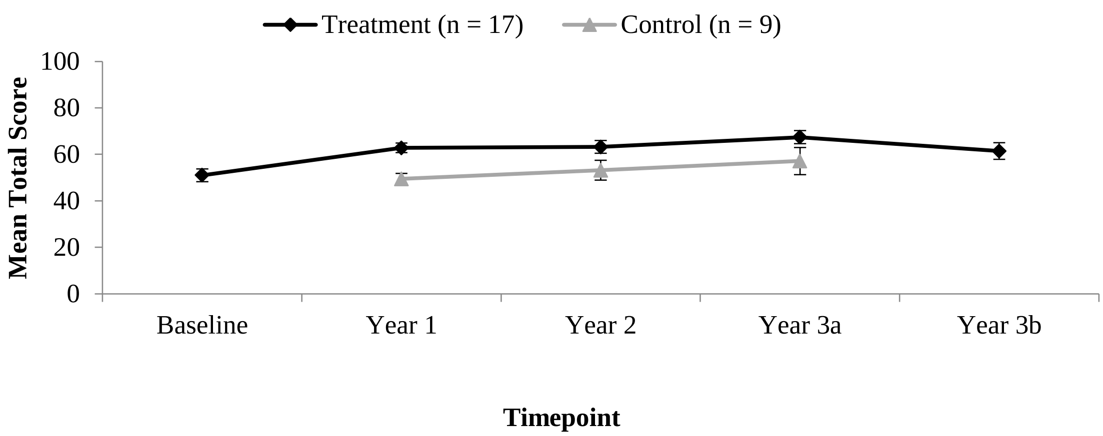

Figure 2. RTOP treatment vs. control comparison (RTOP overall)

In addition to computing overall RTOP scores for treatment teachers, each RTOP subcategory was scored to allow closer examination of how the INSPIRES innovation impacted longitudinal pedagogical growth (Table 1). In general, baseline, SL1 and SL2 scores were relatively low and similar in value, while scores for transfer lessons and, especially EL1 and EL2, were comparatively higher. These trends suggest that treatment teachers did not initially utilize strong reform pedagogies, and instead such pedagogies were most evident during the enactment of engineering design lessons. The subsequent rmANOVA indicated significant differences across the nine lessons for the overall RTOP (25 items) as well as for subcategories of Lesson Design, Procedural Knowledge, Classroom Culture, and Teacher-Student Relationships (Table 2). Teachers' content domain (biology vs. technology education) did not have a significant impact on the pattern of performance (Table 2).

Post-hoc pairwise comparisons revealed that overall average performance on lessons T1, T3a, EL1, and EL2, scored significantly higher than on baseline lessons (Table 3). Average scores for EL2 were significantly higher than all other lessons except for EL1 and T3a. When content domains are examined individually, we only see significantly higher RTOP scores in EL2 when compared to both SL1 and baseline lessons. This may be attributed, in part, to decreased power from sample size. Overall, biology and technology education teachers exhibit the same trends in significant differences among pairwise comparisons of lessons (Table 3).

Pairwise comparisons within RTOP subcategories showed that both EL1 and EL2 yielded significantly higher scores than either SL1 or baseline lessons, in all subcategories except for Propositional Knowledge (Table 4). Scores for EL2 were also significantly higher than those for SL2, T2, and T3b in the area of Lesson Design. An examination of transfer lessons found that T2, T3a, and T3b yielded significantly higher scores than baseline lessons in the subcategories of Procedural Knowledge and Classroom Culture; significant growth in Classroom Culture was also evident by the T1 time point.

Table 1. Mean (SD) scores for RTOP overall and subcategories

RTOP Categories	Baseline	SL1	EL1	T1	SL2	EL2	T2	T3a	T3b
All Teachers (N = 17)									
Overall[a]	49.8 (12.3)	53.6 (12.3)	66.8 (11.4)	63.8 (8.9)	60.5 (10.5)	74.8 (10.8)	64.9 (11.4)	67.4 (11.8)	61.4 (14.7)
Lesson Design[b]	9.5 (3.9)	10.2 (3.3)	13.9 (2.5)	12.9 (3.4)	11.6 (3.0)	15.6 (2.5)	12.9 (3.2)	14.0 (3.3)	11.7 (3.8)
Propositional Knowledge[b]	13.0 (2.2)	13.6 (3.4)	14.1 (3.6)	14.4 (2.0)	14.9 (2.1)	15.2 (2.6)	14.4 (2.8)	15.4 (2.8)	14.9 (3.2)
Procedural Knowledge[b]	8.8 (3.0)	10.6 (1.9)	13.5 (2.0)	12.0 (2.3)	11.9 (2.3)	14.9 (2.4)	12.9 (2.5)	12.6 (2.6)	11.7 (2.9)
Classroom Culture[b]	8.5 (2.8)	9.2 (2.3)	12.4 (2.5)	12.0 (2.2)	11.1 (1.8)	14.5 (2.4)	12.6 (2.5)	12.8 (2.7)	11.5 (3.1)
Teacher-Student Relationships[b]	9.9 (3.2)	9.9 (2.5)	12.8 (2.2)	12.5 (2.3)	11.0 (3.1)	14.5 (2.4)	12.1 (2.7)	12.6 (2.2)	11.6 (3.0)
Biology Teachers (N = 6)									
Overall[a]	45.7 (9.2)	57.8 (6.4)	68.5 (3.9)	63.0 (6.2)	68.2 (10.1)	78.0 (5.1)	66.3 (10.7)	68.7 (11.3)	61.5 (11.3)
Lesson Design[b]	8.3 (3.3)	11.0 (1.9)	14.3 (1.2)	13.3 (2.6)	13.3 (2.8)	16.2 (1.2)	12.5 (4.0)	13.8 (3.4)	11.2 (3.3)
Propositional Knowledge[b]	13.0 (2.4)	15.2 (1.2)	14.3 (1.2)	13.5 (1.6)	16.7 (1.5)	16.3 (1.6)	15.5 (1.8)	16.8 (1.9)	15.7 (2.0)
Procedural Knowledge[b]	7.8 (2.1)	11.3 (2.1)	13.5 (0.8)	11.8 (3.1)	13.2 (2.9)	15.7 (0.8)	13.0 (2.3)	13.0 (2.5)	12.2 (2.9)
Classroom Culture[b]	7.5 (2.0)	9.5 (1.2)	12.5 (1.6)	12.0 (2.2)	12.2 (1.8)	14.3 (2.0)	13.3 (2.5)	12.5 (3.0)	11.2 (1.7)
Teacher-Student Relationships[b]	9.0 (2.1)	10.8 (2.1)	13.8 (1.2)	12.3 (2.1)	12.8 (3.4)	15.5 (1.4)	12.0 (1.9)	12.5 (2.3)	11.3 (2.3)
Tech. Ed. Teachers (N = 11)									
Overall[a]	52.0 (13.6)	51.4 (14.3)	65.8 (14.1)	64.2 (10.3)	56.3 (8.5)	73.0 (12.8)	64.1 (12.2)	66.7 (12.6)	61.4 (16.8)
Lesson Design[b]	10.1 (4.3)	9.7 (3.9)	13.7 (3.0)	12.7 (3.8)	10.6 (2.7)	15.4 (3.0)	13.1 (2.9)	14.1 (3.4)	12.0 (4.2)
Propositional Knowledge[b]	13.0 (2.1)	12.8 (4.0)	14.0 (4.4)	14.8 (2.0)	13.9 (1.6)	14.5 (2.9)	13.8 (3.1)	14.5 (2.9)	14.5 (3.7)
Procedural Knowledge[b]	9.4 (3.4)	10.3 (1.8)	13.5 (2.4)	12.1 (1.9)	11.2 (1.7)	14.5 (2.9)	12.8 (2.7)	12.5 (2.8)	11.5 (3.0)
Classroom Culture[b]	9.1 (3.1)	9.1 (2.8)	12.4 (2.9)	12.0 (2.3)	10.5 (1.6)	14.6 (2.7)	12.2 (2.5)	13.0 (2.6)	11.7 (3.7)
Teacher-Student Relationships[b]	10.5 (3.7)	9.5 (2.6)	12.3 (2.5)	12.5 (2.6)	10.0 (2.5)	14.0 (2.7)	12.2 (3.1)	12.6 (2.3)	11.7 (3.5)

[a]100 points possible. [b]20 points possible.

Table 2. Repeated Measures ANOVA of RTOP Scores Across Lessons

	Categories:	**RTOP Overall**	**Lesson Design**	**Propositional Knowledge**	**Procedural Knowledge**	**Classroom Culture**	**Teacher-Student Relationships**
Lesson Comparisons	**F(2,50)**	11.037***	8.049***	2.054	12.101***	11.413***	7.620***
Biology vs. Tech. Ed.	**F(2,50)**	1.207	0.865	1.223	0.946	0.742	1.486

*$p < .05$, **$p < .01$, ***$p < .001$.

International Journal of Research in Education and Science
Volume 5, Issue 2, Summer 2019 ISSN: 2148-9955

Table 3. Pairwise comparisons: RTOP Overall and by Content Area

Comparison	Sig. (All Teachers N = 17)	Sig. (Bio Teachers N = 6)	Sig. (Tech Teachers N = 11)
EL1 : Baseline	**<.001***	.088	.087
T1 : Baseline	**.049***	1.000	.822
EL2 : Baseline	**<.001***	**.022***	**.017***
T3a : Baseline	**.008***	.963	.195
EL1 : SL1	**.016***	.466	.123
EL2 : SL1	**.001***	**.031***	**.045***
EL2 : T1	**.030***	.209	1.000
EL2 : SL2	**.046***	1.000	.117
EL2 : T2	**.018***	.937	.338
EL2 : T3b	**.031***	.183	1.000

Note: Within each comparison, the first lesson named indicates the lesson with the higher average RTOP score. Only pairwise comparisons yielding significant differences (Tukey's HSD) are shown.

Table 4. Significant pairwise comparisons for RTOP Subcategories

Comparison	Lesson Design	Propositional Knowledge	Procedural Knowledge	Classroom Culture	Teacher-Student Relationships
EL1 : Baseline	**.009***	1.000	**<.001***	**.005***	**.006***
EL2 : Baseline	**.001***	.085	**<.001***	**<.001***	**.002***
EL1 : SL1	**.011***	1.000	**.023***	**.010***	**.007***
EL2: SL1	**.002***	1.000	**<.001***	**<.001***	**<.001***
EL2 : SL2	**.010***	1.000	.123	**.019***	.055
EL2: T2	**.038***	1.000	.074	.455	.112
EL2 : T3b	**.018***	1.000	**.012***	**.021***	**.042***
T2 : Baseline	1.000	1.000	**.001***	**.016***	1.000
T3a : Baseline	.093	.073	**.003***	**.029***	.177
T3b : Baseline	1.000	1.000	**.014***	**.039***	1.000
EL2 : T3a	.853	1.000	**.027***	1.000	.126
T1 : Baseline	.266	1.000	.161	**.029***	.521

Note: Within each comparison, the first lesson named indicates the lesson with the higher average RTOP score. Only pairwise comparisons yielding significant differences (Tukey's HSD) are shown.

Qualitative Results

Qualitative results were determined by identifying common themes from a subsample of six teacher participants; three biology and three technology education teachers. This subsample of teachers produced lessons with RTOP scores around the means representative of all participating teachers, and thus characterize common trends in teaching practices throughout the study. Baseline recordings were meant to capture teachers' best attempt at a design-based lesson prior to receiving any intervention through the INSPIRES PD and educative curriculum project. Teachers A, B, and D taught biology and led lessons around identifying local endangered species, creating a biological 'super hero,' and building a solar oven, respectively. Teachers C, E, and F taught technology education and led lessons on designing and building propellers, a heart-lung machine, and Lego cars, respectively. Among the six teachers sampled for qualitative review, four teachers assigned traditional bell work to students at the beginning of class (Teachers A, B, C, and D). Teachers A, B and C also spent the early part of the class period lecturing to students, especially regarding the introduction of vocabulary. All six teachers incorporated a hands-on activity for students; students worked in pairs or small groups in all cases except for Teacher C. Most often, the activities had step-by step instructions and were not considered open-ended in nature (Teachers A, B, D, E). In some cases, the teachers were more likely to direct students or provide answers to questions in lieu of wait time or probing for student rationale (Teachers A, C, D). Most cases did not include a share-out, recap, or other debrief at the end of the lesson (Teachers A, B, C, D). In no case were design-based decisions explicitly linked to underpinning scientific rationale; only Teacher E explicitly linked mathematic rationale to students' design decisions. Teacher E's baseline was unique in that he enacted a lesson that was previously developed by the same research team as the current INSPIRES study, which was part of an engineering design challenge within an educative curriculum (to design, build and test a model heart lung machine). Teacher E made reference to the engineering design loop (process), expected students to use mathematical rationale, and asked students to write reflections in their engineering design journals based on 'best next steps in design considerations,' for homework. The lesson of Teacher F was also a unique baseline measurement since he opened the lesson by expecting students to identify the problem, criteria and constraints of an engineering design challenge (of building a Lego car). Teacher F shared relevant student ideas with the whole class, built in opportunities for redesign, and was the only teacher in the qualitative review to include a whole-group reflection at the end of the class period.

During Year 1 implementation of the INSPIRES curriculum, several themes emerged from how teachers generally enacted the focal science lesson (SL1). First, although the SL1 lesson plan encourages use of artifact sharing in lieu of traditional bell work, five of six teachers used traditional drills to begin the lesson. It is worth noting, that Teachers A and C prompted artifact sharing following traditional bell work; Teacher E also asked students to share artifacts yet students were not prepared and did not participate. Teacher F was the only case where student driven artifact presentations and KWL chart reflection fully replaced a traditional drill. A second dominate theme was the commonality of teachers assuming control over parts of SL1 that were originally designed to allow for student autonomy. For example, students in all six classrooms were assigned pre-conceived experimental variables and provided with pre-determined protocols. It was also common for teachers to contribute the bulk of conceptual explaining (Teachers A, B, C) and even calculate resulting concentrations of waste removal for their students (Teachers A, C). None of the enactments included a whole class reflection on the day's lesson and its connection to the overall design challenge; Teachers B and D did prompt their students to individually consider how their results influence the design of their hemodialysis systems. Several teachers prompted their students to sketch and label their experiments (Teachers B, C, F), or make other notes (Teachers D, E), within their engineering design notebooks. Multiple teachers prompted students to think about or record hypotheses (Teachers A, B, F), yet whole-class discussions of hypotheses were absent. Teachers A and B made reference to the engineering design loop (process) and design target (criteria/constraints), yet drove these discussions themselves. Teacher C was the only case that followed the SL1 lesson plan suggestion to provide a tangible example of diffusion, however, this was presented after the frontloading of terms and definitions and therefore lacked the 'phenomena first' impact intended by the lesson plan. Teachers B, C and E made implicit connections between SL1 and a prior lesson in the unit, that allowed students to think about filtering particles at the macro level.

In general, the six focal teachers of the qualitative review ran an open-ended and student-driven enactment of INSPIRES EL1, as intended by the lesson plan. Half the teachers arranged for their students to present artifacts (Teachers A, D, F) while another teacher (E) used a KWL chart to record his students' ideas and questions. Most teachers (A, B, E, F) referenced prior lessons in the INSPIRES Hemodialysis unit and made connections between collected data and pending design decisions. Reference to the engineering design loop (process) and the design target (criteria and constraints) was evident in the classrooms of Teachers B, C, D, E, and F. Many Teachers (B, C, E, F) also encouraged student autonomy by allowing students to bring in their own building materials from home, and explicitly conveying the expectation of divergent solutions across student groups. In particular, Teachers A and C mentioned that groups could choose whether or not to use pumps in their designs, while Teacher E showed photos to emphasize the variety of designs used in the past. Nearly all teachers (A, B, C, D, E) expected students to use their engineering journals to reference information, sketch designs, and/or reflect on connections or next steps. Both Teachers B and F frequently pressed students for scientific understanding and rationale for design decisions; the other four teachers either directed scientific discussions themselves or did not highlight science content as a focus of the lesson. Most teachers did not incorporate a whole-class debrief at the end of the lesson. However, Teachers A and D prompted sharing of plans (without discussion or critique) from a few student groups, while Teachers E and F expected students to individually reflect on the science of their best solutions through a writing assignment.

Teacher-selected lessons following the first enactment of the INSPIRES curriculum, T1, covered a wide variety of topics with open-ended components. Biology Teachers B and D led the same general lessons as during the baseline; creating a biological super hero and designing and building a solar oven, respectively. Teacher A shared a lesson on taxis behavior in ladybugs. Technology education teachers led lessons focused on Bernoulli's Principle and wing design (C), designing and testing mousetrap cars (E), and using a simulation to explore AC/DC circuits (F). All six cases included activities that had design, build, test, or other exploratory components; five of six cases also had students working in pairs or small groups (Teachers A, B, D, E, F). Most lessons began with traditional bell work that prompted students to write and share definitions or terminology (Teachers A, B, D, F); Teacher C used a phenomena-first exercise followed by questioning the whole class and Teacher E asked his groups to begin by measuring the diameter of their car wheels and sharing this information with the whole class. Half of the T1 lessons could be described as design challenges, yet the connections to underlying scientific principles were generally absent. It was also common for teachers to place restrictions on what could have been student-led components of the lesson. For example, Teacher A discouraged a student from testing visual taxis in ladybugs (instead of chemotaxis), Teacher C expected students' wing designs to be symmetrical and have a similar "thick to thin" shape as his example, and Teacher E expected students to test their mousetrap cars on a predetermined timeline (possibly before the students were ready). At the end of the lessons, a few classes reported out on what each student or group was exploring, but without inviting discussion or critique (Teachers A and B). Alternatively, Teacher E expects his students to individually document three potential design modifications in their engineering journals before leaving class. The remaining three lessons did not include student reflection or sharing out at the conclusion of the period. Teachers A and E expected students to form hypotheses or predictions, yet scientific rationale is not emphasized. Students were also expected to have their respective designs checked and approved by Teachers C and D, prior to building with materials. While most teachers facilitated progress of individual students or student groups throughout lessons (Teachers B, C, D, E, F), only Teachers B and F pressed students to communicate scientific rationale.

Overall, enactment of SL2 was very teacher-directed and pedagogies were similar to SL1. Four of six teachers continued to use traditional bell work, typically around terminology definitions, in lieu of the artifact presentation suggested by the SL lesson plan (Teachers C, D, E, F). Of the two classrooms that used artifacts, one only sometimes made explicit connections to scientific concepts (A) and the other involved a single teacher-presented artifact (C). In each of the six cases, the teacher had pre-selected the variables to be tested and given the full protocols and pre-organized materials to student groups. One teacher (D) did allow her students to choose which if the pre-determined variables they were interested in testing and form groups based on that preference. Some teachers expected students to use their engineering notebooks to sketch experimental set-ups and record data (A, B, E, F), while other teachers seemed to have students record pertinent information on a worksheet (C, D). Half the sampled teachers (A, D, F) expected students to pose hypotheses or make predictions during the lesson and two of these teachers pressed students further for scientific rationale (A, F). Most teachers (A, B, C, F) refer to the overall design target (criteria and constraints) during the lesson; two of those teachers (A, F) make an explicit connection with how control conditions of SL2 do not meet the target and other factors need to be considered for the broader hemodialysis system design. Several teachers (B, D, E, F) also allowed their students to compare their experiment samples to the reference standard and carryout the mathematical computation to determine their resulting concentration of waste. Three of the sampled classrooms (B, C, F) had whole-group data charts or other means of sharing results between student groups testing different variables; yet none of the classes included a reflective conclusion to connect students' findings with the greater design challenge. Several teachers maintained control over one or more aspects of SL2 that was intended to be student-centered. In particular, Teachers A and C calculated the resulting waste concentration for their student groups; Teachers E and F pipet fluid volumes for all student groups; Teachers A, B, and D would occasionally give away answers or imply correct hypotheses before the students had a chance to discover the information for themselves through the activity. A few of the biology teachers (A, B) led explicit discussions on the underpinning scientific concepts of diffusion and equilibrium, while a few of the technology education teachers indicated less comfort with such concepts by asking the audio-recorder how to measure waste with the standard (C) and using vague terminology and descriptions when working with students investigating the impact of varying concentration gradients on diffusion (E).

All six teachers began EL2 with drills that prompted students to think about what they had recently learned from a prior INSPIRES lesson, although only two teachers used the format of artifact presentations (Teachers A and C). Additionally, Teacher B, E, and F had individual students discuss their ideas with a small group before sharing responses with the whole class. All six teachers either discussed the design target or individual criteria and constraints with their classes, yet only Teachers B and E referenced where the class was within the design loop. Most Teachers (A, B, E, F) pressed students for scientific and/or mathematic rationale as students described their design ideas for their Teacher's approval. All teachers expected their students to sketch design ideas prior to building, and a few teachers prompted students to include these sketches as part of an ongoing engineering design notebook (A and E). Teachers A and B encouraged their students to bring materials from home to incorporate into their final designs. Most teachers prompted their students to consider how information from previous lessons could inform design decisions. For example, Teachers B and F asked students to think about scientific concepts of SL2, Teachers A, B, C, and F asked students to revisit information from a simulation lesson, and Teacher E asked students to consider the pros and cons to using various pumps that students had explored in the prior class. Both Teachers D and E also shared pictures of their personal designs as examples to promote divergent thinking and to give students a starting frame of reference. While Teachers C and E did not explicitly discuss underpinning science concepts themselves, they did take time to replay videos and simulations from prior INSPIRES lessons that included scientific demonstrations. Teacher B exhibited several standout student-centered traits including flexible grading (i.e., she would regrade a students' sketches if they continued to add details, such as rationale), explicitly stated that students would create their own procedures, and rephrased questions when students were confused, rather than giving away answers. Teacher F also drew sketches or used other tangible examples to help make abstract scientific and math concepts more concrete when students showed signs of confusion. None of the classes incorporated an end-of-period debrief, with the exception of Teacher F who asked groups to write down their next steps and struggles on post-it notes that were displayed in the classroom.

Transfer lessons in Year 2 covered a variety of different STEM topics. Biology teachers led lessons on ladybug kinesis behavior (A), the evolution of Galapagos finches' beaks (B), and the design and building of solar ovens (D). Technology Ed. teachers shared their lessons on designing and building propellers from sheet metal (C), refining and testing the performance of mouse trap cars (E), and researching various construction methods prior to designing and building a physical model (F). Several teachers (A, B, C, D) began their lessons with traditional bell work as students individually responded to knowledge-based questions or vocabulary review. Conversely, Teachers E and F opened their lessons with student reflections on challenges/approaches and student contributions to a KWL chart, respectively. Only two teachers (A, C) spend early class time on the introduction of relevant vocabulary, while several teachers (A, B, C, D) used videos to introduce concepts related to the day's activity. Most of the lessons had open-ended components (Teachers A, C, D, E, F). Further, Teacher A expected his students to plan their own procedures, sample sizes, and hypotheses related to their ladybug kinesis experiments. While Teacher B allowed students to contribute various scientific ideas and questions toward the beginning of the class, the main activity was not tightly connected to scientific concepts, as was the case in several of classrooms (Teachers A, C, D). All six lessons included a hands-on component that kept students engaged; half of the lessons included small group-work (B, D, E). A few of the Technology Ed. teachers made specific reference to mathematical concepts as they prompted students to consider radial degrees of their propeller designs (Teacher C) and diameters of their car wheels (Teacher E). Teachers D, E, and F all expected students to provide labeled sketches or other documented plans prior to building and testing. Half of the teachers did not prompt an end-of-class wrap-up, while the wrap ups varied among the other half. Specifically, Teacher D asked small groups to report out plans to overcome challenges, Teacher E prompted students to individually reflect on their day's accomplishments in their journals, and Teacher F returned to the KWL chart for a whole class discussion.

Many of the Transfer 3a lessons included open-ended design components and covered different topics than previous transfer lessons. Biology lessons included a cell diffusion lab (A; no open-ended component), designing a prosthetic leg, and designing spacecraft qualities to withstand freezing temperatures. Technology Ed. lessons included designing leak-proof connectors for tubing (C), similarly designing leak-proof adapters for an IV bag (E), and designing a crane (F). All six lessons began with students completing traditional bell work related to defining terminology, homework review, or other knowledge-based questions. Both Teachers B and D allowed flexible timing for students to complete drills, and Teacher C followed up the drill with an exercise that asked pairs of students to reflect on how a finger trap toy works. Several teachers provided sketches, demonstrations, or videos to aid their students in grasping abstract concepts or problems (Teachers, A, B, D, E). Almost all students were working in small groups (Teachers A, B, D, E, F) and several teachers (B, D, E, F) expected students to sketch their designs both individually and as a group. Further, Teachers B and D required that student groups seek teacher-approval of final designs before they begin building products. Teachers B and C referenced the engineering design loops while Teachers B, D, E and F either implicitly or explicitly referenced project criteria and constraints. Teachers A and B guided their students in data collection and interpretation as critical steps in drawing conclusions. Most of the teachers made explicit connections to math concepts within their lessons. For example, Teacher B expected her students to measure stride lengths and compute averages before determining specifications for a prosthetic leg. Teachers C and E emphasized the importance of using diameter measurements to inform the design of tubing connectors and adaptors. Teacher F also expected his students' crane design sketches to reflect mathematical proportions. Half the teachers allowed their students to contribute the majority of the ideas and rationale for design decisions (B, E, F), while Teacher D expected her students to provide explanations some of the time. A few teachers made efforts to engage students through use of popular media, as Teacher B played a movie scene of a zombie attack (to set the background story for why students needed to design prosthetic legs), while Teacher F played a song from a popular movie at the beginning of his class. Only two teachers prompted students in an end-of-class wrap-up; Teacher D asked groups to collectively fill out a KWL chart and Teacher E asked individual students to write journal reflections on changes made to CAD designs, including rationale.

For Transfer lesson 3b, half the subset of teachers led open-ended activities. Biology Teacher B asked her students to research and pose what they found to be important components of a healthy coral reef; Technology Ed. Teacher E guided his students in using a computer simulation to explore the various factors that contribute to successful bridge design. Teacher F had a unique lesson plan that asked students to consider finding an innovative solution to a real world problem of a community stakeholder. A traditional, confirmational lab was led by Teacher A, in which students followed a procedure to separate salt, sand, and iron from water. Teacher C provided step-by-step instructions for his students to design a specific tubing connector with CAD software. A biomedical PLTW unit was the focus of Teacher D's class, where students interpreted graphical information to draw conclusions about a diabetes diagnosis. While all six classes began with traditional bell work, both Teachers E and F gave multiple students opportunities to share their responses and ideas with the whole class. Several of the lessons included quantitative components such as making measurements in length or weight (Teachers A, C, E), computing averages (A), or making meaning from data tables and graphs (D). Half the lessons prompted students to carefully record their ideas, observations, design sketches, or other findings (Teachers A, E, F).

Most of the lessons included a time when students were expected to work with a partner or small groups (A, D, E, F). Teachers B, D, and E frequently expected students to provide explanations and scientific rationale related to their assignments; Teacher F also pressed students for rationale but more related to students' personal capabilities to solve a community problem rather than explicit ties to scientific concepts. Nearly all lessons had practical relationships with real world scenarios (B, D, E, F). For example, Teacher D asked her students to comment on their knowledge of diabetes and related lifestyle challenges in their community. End-of-class debriefs across the whole group of students was evident in five of six lessons (A, B, C, E, F); alternatively, Teacher B had individual students write reflections. Such debriefs varied between teachers sharing ideal outcomes or next steps (Teachers A, C) to students sharing ideas of proposed solutions or lessons learned (Teachers E, F).

Interpretations

The present study evaluated teachers' pedagogical changes over the course of a three-year PD program, as a function of their participation in the combined INSPIRES educative curriculum and PD. Overall, growth in reformed pedagogy was evident between baseline and end-of-year-three lessons when measured with the RTOP rubric. A closer examination of specific RTOP subcategories reveals that significant pedagogical growth occurred in the areas of Classroom Culture and Procedural Knowledge, but not as much in Propositional Knowledge (Table 4). This discussion begins by addressing the results for RTOP subscales, followed by the close consideration of INSPIRES lesson plans and INSPIRES as an educative curriculum. The discussion reflects the role of teachers' STEM content areas, and the value of control group comparisons, among the findings. Finally, we discuss the limitations, implications, and next steps for the study.

Classroom Culture

The other RTOP subscale of notable growth, Classroom Culture, is measured by items that assess the degree of students' active communication, ability to ask questions, dominate the critical discussions, ideas receive careful consideration, and how the teacher may further facilitate students' divergent thinking. Unlike, Procedural Knowledge, significant growth in Classroom Culture was evident as early as T1. Baseline lessons tended to include teacher-dominated discussions, including lecturing and the asking and answering of questions, and only one teacher in six created an opportunity for students to communicate their ideas to each other. After one INSPIRES enactment and one year of supporting PD, it is already more common that teachers allow students to communicate their ideas through technical sketches, explain scientific rationale for design decisions, and share thoughts with students outside of their small groups. Further, student communication may have been enhanced by more lessons structured to encourage students to work in small groups or pairs.

Procedural Knowledge

Procedural Knowledge is scored through the RTOP with items that assess the extent to which students are allowed to direct the process and progression of the lesson. In particular, classrooms that encourage multiple means of expression, student driven hypotheses, student driven procedures, active reflection, and rigorous student discourse are classrooms well-suited to score highly on the Procedural Knowledge RTOP subscale. Significant growth in Procedural Knowledge from baseline levels began with lesson T2 ("transfer lesson for Year 2) and continued through lessons T3a and T3b (Transfer Lessons a and b for Year 3). By lesson T2, teacher participants had enacted the INSPIRES curriculum twice, which provided a model for how lessons could be open-ended and student-centered. While INSPIRES science lessons SL1 and SL2 (INSPIRES Lesson 7, which had a heavy emphasis on Science concepts, for Years 1 and 2) did not show significant growth from baseline levels of Procedural Knowledge, INSPIRES engineering lessons EL1 and EL2 did show such growth. In Baseline lessons, while most students engaged in a "hands on" activity, most of these activities could not be considered open-ended because the teachers scripted the activities so that every group achieved the same pre-determined result. Further, most Baseline lessons included teacher-directed lectures or giving-away of information prior to student exploration. So the exploration became more of an enrichment and reinforcement activity rather than the intended phenomenon-driven learning. There was no mention of student-driven hypotheses or press of student-driven scientific rationale; in most cases there was no active reflection. In other words, Baseline lessons very commonly yielded qualities that signified them as teacher-directed, which explains the relatively low RTOP Procedural Knowledge scores.

Alternatively, lessons T2, T3a and T3b (transfer lessons—teacher-made lessons in Years 2 and 3) included traits that suggest the lesson process and progression has become more student-centered. In particular, qualitative findings showed that these later lessons were more likely to begin and end with student reflections, and many contained open-ended components that required students to derive hypotheses, procedures, and evaluate outcomes. Those lessons that included the designing and building of a project often required students to document plans (i.e., labeled sketches) and seek teacher approval prior to construction. These reformed qualities are also highlighted throughout the INSPIRES educative curriculum. The academic year PD sessions during Year 2 emphasized the function of engineering design challenges and the engineering design loop. Teachers tended to include reformed practices in lessons EL1 and EL2 (engineering) more than SL1 and SL2 (science), suggesting that engineering projects may be structured to support reformed practices more than stand-alone science lessons.

Propositional Knowledge

The INSPIRES engineering lessons, EL1 and EL2, were the standout cases for pedagogical reform. Pairwise comparisons within RTOP subcategories showed that both EL1 and EL2 yielded significantly higher scores than either SL1 or baseline lessons, in all subcategories except for Propositional Knowledge (Table 4). Propositional Knowledge is measured by assessing the degree to which the teacher has comfort and expertise in the content area. In relationship to pedagogical reform, teachers with a strong foundation in content knowledge are among those that feel more comfortable releasing control of the lesson to students, and allow student ideas and questions to direct the scope of the lesson (Sawada et al., 2002). In other words, the teacher has expertise that prepares them to respond to, and facilitate conversations based in divergent student ideas. Lessons EL1 and EL2 likely did not significantly outshine Baseline and Transfer lessons in the area of Propositional Knowledge because teachers probably self-selected the latter lessons in part, on the basis of high comfort level. This reasoning could also explain why there is not significant growth in Propositional Knowledge between Baseline and Transfer lessons; at each time point, teachers had the freedom to select which of their lesson repertoire they wanted to have recorded and critiqued.

Lesson Design

Scores for EL2 were also significantly higher than those for SL2, T2, and T3b in the area of Lesson Design. The INSPIRES lesson plans for both SL and EL contain most of the pedagogical elements highlighted within the RTOP, however, EL arguably contains more than SL. For example, within Lesson Design, the EL lesson plan explicitly encourages the reformed elements of RTOP Items 1-5 while the SL lesson plan explicitly encourages only RTOP Items 1-4. Item 5 deals specifically with whether the focus and direction of the lesson is determined by students. In the INSPIRES SL lesson plan, the teacher acts to guide the lesson so that multiple factors impacting diffusion emerge by the end of the period. Student prior knowledge, ideas, communication and autonomy (such as identifying factor to test) are all encouraged by the SL lesson plan, but are evaluated by other items on the RTOP. The lesson plan for EL explicitly addresses RTOP Item 5 by encouraging student design groups to work at their own pace through the design cycle. In other words, EL is actually a multiple-period lesson in which students design, build, test, redesign and ideally retest their hemodialysis systems. Since some students may be finishing their design while other students may be testing, this flexible timing is planned so that students may direct the focus of the lesson. The fact that EL2 was significantly more reformed than T2 and T3b, but not T1 and T3a, in the area of Lesson Design suggests that teachers may not be making consistent changes to shift the design of their lessons over time. When examining the topics and structure of each transfer lesson from the subset of teachers for qualitative review, 5-6 teachers incorporated open-ended components with T1 and T3a, while 3-4 teachers did the same for T2 and T3.

INSPIRES Curriculum Mapped to RTOP

When mapping the RTOP to the focal science and engineering INSPIRES lessons as written, it seems that in addition to Item 5, Items 15, 19, and 25 were explicitly encouraged in the EL1/EL2 lesson plan but not the SL1/SL2 lesson plan. This difference in lesson design of the curriculum may also play a role in teachers' tendency to exhibit more reformed qualities during EL1/EL2 lessons than SL1/SL2 lessons. For example, Item 19 of the RTOP falls under the Classroom Culture subscale and focuses on whether student comments or questions determine the direction of classroom discourse. While the SL1/SL2 lesson plans call for discussions around student ideas at the beginning (i.e. artifact sharing) and end (i.e. whole class reflection/debrief) of the period, the lesson plan does not explicitly prompt teachers to allow students to lead the conversations. Instead, the conclusion tips for SL1/SL2 are more structured and even suggest that teachers ask some close-ended questions of students. Alternatively, the lesson plan for EL1/EL2 frames the lesson conclusion as expecting student groups to share their progress with the whole class including what data informed their design decisions, what are their next steps, and what are their obstacles. Such a discussion poses a student-led culture and is open-ended in direction. RTOP Item 25 lies within the Teacher-Student Relationships subscale and asks the degree to which teachers act as listeners and support students in learning gains crafted from their pre-understandings. The lesson plan for EL1/EL2 explicitly directs teachers to provide feedback as needed while monitoring student groups and to probe for STEM-based rationale for design decisions. Alternatively, the lesson plan for SL1/SL2 makes no explicit mention of group-monitoring or pressing for rationale, yet, these reformed pedagogies may have been implied in other ways. Finally, RTOP Item 15 deals with "intellectual rigor, constructive criticism, and the challenging of ideas" (Piburn & Sawada, 2000) which is central to the EL1/EL2 lesson plan as teachers facilitate student-led design by checking and questioning student designs and probing for STEM-based rationale. While the lesson plan for SL1/SL2 asks teachers to solicit student ideas for testable variables and related procedures, there are fewer explicit prompts that encourage teachers to challenge student ideas. Likewise, among the subsample of six teachers for qualitative review, all teachers provided students with pre-selected variables and pre-determined protocols for both SL1 and SL2, which hindered the opportunity for students to contribute ideas. Although some of these teachers did expect students to record hypotheses during SL1 and SL2, these students' ideas were largely left unchallenged or not revisited later in the period. Alternatively, students were frequently pressed for STEM-based rationale of design decisions by two teachers during EL1 and by four teachers during EL2. In summary, the lesson plan for EL1/EL2, as written, is relatively stronger in the practice of argumentation, than the lesson plan for SL1/EL2.

Argumentation is of particular importance to the development of scientific discourse because it invites socially constructed understanding of claims based on evidence, a practice valued by the NGSS. Although teachers scored relatively higher on Item 15 during EL1/EL2 than on SL1/SL2, mean Item 15 scores were the lowest among all 25 items rated during both EL1 and EL2. This relative weakness on Item 15 suggests that pedagogical strategies to promote argumentation may not be explicitly conveyed in the INSPIRES EL1/EL2 lesson plan. The EL1/EL2 lesson plan encourages teachers to direct students to work as teams to finalize a "best idea" for the design challenge; however, pedagogical tips to suggest argumentation as a means for finalizing a best idea, were lacking. The SL1/SL2 lesson plan proposes that students share their hypotheses/predictions with the whole class and work in small groups to develop their own procedures, but there is no mention of argumentation as a strategy for making deeper meaning within these class discussions. Several of the focal teachers included in the qualitative review would press students for STEM-based rationale in at least one lesson, yet a framework or expectation for students to consistently critique ideas was missing. A post hoc look at how teachers faired specifically in SL and EL lessons revealed that item 15 scores are among the lowest of any item on the RTOP for those lessons (data not shown). Since argumentation is a pedagogical skill with multiple components around discourse (i.e., Ask, Press, Link; e.g., Fishman, Borko, Osborne, Gomez, Rafanelli et al. 2017), a teacher might only incorporate it into the INSPIRES lessons (or other lessons, for that matter) if the teacher was already familiar and practiced in the strategy. Indeed, argumentation is largely absent from K-12 classrooms (Osborne 2010) and teachers require more support for the development and transfer of argumentation skills into their own classrooms (Simon, Erduran, & Osborne, 2006). McNeill, González-Howard, Katsh-Singer, and Loper (2017) identified three critical components influencing teachers' tendency to adopt argumentation pedagogy following related professional development: 1) teachers' understanding of argumentation as a valid and valuable practice, 2) teachers' willingness to reflect and critique curricula, and 3) teachers' discontent with their past pedagogies or teaching experiences. Similarly, teachers benefit from opportunities for active reflection when learning new pedagogies, particularly in the context of argumentation (Marco-Bujosa et al. 2017). There is currently a need to further support teachers in the adoption of scientific argumentation, both philosophically and in practice, as the NGSS become fully realized across the nation. While the INSPIRES educative curriculum highlighted in the present study does not explicitly embed extensive support in argumentation, the authors recognize the value of such scientific discourse and will make argumentation the focus of future studies.

Another finding is that while overall RTOP scores for EL1 and EL2 are significantly greater than SL1, they are not significantly greater than SL2. This suggests a trend of growth between the enactment of SL1 and SL2 (although not significant with respect to each other) which is in alignment with teachers general sentiment that the second year of the INSPIRES curriculum felt more successful than the first year of enactment (expressed during multiple PD sessions; data not shown). Further, consideration of the qualitative features of these lessons demonstrates that an increased proportion of teachers provided explicit discussions of related scientific topics (i.e., diffusion, concentration gradient, etc.), expected student groups to share and record outcomes in a class data table, and expected students to make their own comparisons and calculations of results. It is reassuring that a trend of growth occurred between the first and second years of implementation of both the science and engineering lessons, respectively. It is common for significant and sustained pedagogical growth to required more frequent and prolonged professional development than even the three consecutive summers (plus spanning academic year sessions) provided through the present INSPIRES program (Banilower, Heck, & Weiss, 2007; Boyle, Lamprianou, & Boyle, 2005). Even so, the finding of EL1 and EL2 as significantly more reformed than most other focal lessons of the study suggests that a high-quality engineering curriculum lends itself well to the natural incorporation of reformed pedagogical strategies. Namely, the open-ended and problem solving nature of engineering is a rich opportunity for students to drive their own investigations, stemming from their own prior knowledge and socially constructed understandings. A high-quality engineering curriculum will incorporate small group work, explicit ties to science and math concepts, student reflections, a value of divergent thinking, among other qualities. The INSPIRES curriculum was developed ten years ago and has undergone numerous revisions and iterations of enactment and study. Co-authored by engineering and pedagogical experts alike, INSPIRES is arguably among the highest quality curricula for infusing engineering practices and principles into high school classrooms of all types of learners.

STEM Content Area

One of the aims of the present study was to determine if a difference exists in the way that biology and technology education teachers adopt reformed pedagogies when infusing engineering into their lessons. An analysis of all nine time points across the 3-year project indicated no significant differences in RTOP scores between the focal content areas for both overall scores as well as individual subscales. Such similarities were consistent throughout the entire project. After Year 1 (four time points), no difference was detected between biology and technology education RTOP scores (Williams et al., 2019). Similarly, no significant difference was detected between biology and technology education teachers after Year 2 (seven time points; data not shown). The 3-year project witnessed teacher participant attrition over time, which is common among longitudinal studies of educational research. After Year 1, analyses could be completed on 27 participants (Williams et al., 2019) and by the end of Year 3 analyses those same analyses can be applied to the remaining 17 participants. The repeated analyses over multiple time points and varying sample sizes yield consistent patterns of teacher performance by content area. This consistency may be evidence that the present results capture the true nature of pedagogical change, rather than an artifact of limited project duration or reduced statistical power related to sample size or multiple comparisons (of lesson time points).

The absence of any striking difference between biology and technology education teachers was a surprise. We predicted that biology teachers would excel in SL1 and SL2 in comparison to technology educations teachers, because of the former's specific background and expertise in lab-based science lessons. Likewise, we predicted that technology education teachers would excel in EL1 and EL2 in comparison to biology teachers because the former may traditionally include more opportunities to teach design-based lessons. Results indicated, however, that biology and technology teachers did not differ significantly. Because the RTOP measures general elements of pedagogical reform in the context of STEM lessons, the tool can be equally sensitive to pedagogical strategies used in biology, technology education, or engineering lessons. Teachers receiving PD in both pedagogy and the INSPIRES curriculum with an explicit lesson plan to follow can therefore perform at near-equal levels on the RTOP irrespective of the alignment between their teaching background and the precise content of the measured lesson. Williams et al. (2019) discussed various reasons for why biology and technology teachers performed similarly on pedagogical standards across different lessons in Year 1, including: 1) biology teachers may be less likely to follow the INSPIRES lesson plan during SL1/SL2 because of confidence in background/expertise, but may then miss out on subtle pedagogical strengths of the lesson and instead lead a "traditionalized," teacher-led version, 2) technology educational teachers may be more likely to follow the INSPIRES lesson plan during SL1/SL2 because of a lack in background expertise and therefore may incorporate some of the intended pedagogies, 3) the opposite pattern of biology and technology teachers following the EL1/EL2 lesson plan more or less closely, respectively, for reasons similar to (1) and (2), and 4) the overall open-ended nature of INSPIRES EL1/EL2 provided many opportunities for reformed elements of pedagogy, even in the absence of content expertise or precise adherence to the lesson plan. Feedback from teachers during PD sessions and lessons in Years 2 and 3 corroborated these preliminary hypotheses.

For example, when considering how the focal subset of teachers for qualitative analysis lead lessons SL1 and SL2, we find that both biology and technology education teachers assign predetermined variables and provide predetermined protocols to students. The lesson for SL1/SL2, however, encourages teachers to allow students the opportunity to apply their prior knowledge and critical thinking skills, along with teacher guidance, to deduce relevant measurable variables and logical protocols for themselves. When teachers assume more control over these parts of the lesson, there are missed opportunities for students to apply creativity and higher-order thinking skills to solve a problem. Although biology teachers likely have more expertise leading science laboratory lessons and may be predicted to have greater comfort with the associated content, the biology teachers in the present study chose to apply structure to parts of the lesson intended to offer growth through open-endedness. Whether science teachers elected to take control over such parts of the lesson because of discomfort with students assuming control, simply habit of practice of pre-INSPIRES-PD routines, or for some other reason, is not yet understood. As a tool to measure relative degrees of pedagogical reform, the RTOP is sensitive to whether teachers give students opportunities to direct parts of the lesson or use divergent thinking. Therefore, if both biology and technology teachers avoided reformed pedagogical practices, regardless of the reason, then it is logical for both groups to score relatively low on the RTOP for science-based lessons such as SL1/Sl2. In contrast, both biology and technology education teachers in the qualitative focal group encouraged student-driven designs and the divergent thinking that led to a variety of solutions in the engineering-based EL1/EL2 lessons. During INSPIRES PD, the expectation that students would develop multiple, diverse solutions to the hemodialysis problem was made explicit to participating teachers. Further, teachers worked through the lesson in the role of students and witnessed first-hand the variety of systems that could be designed, built, tested, and succeed (meet target criteria and constraints). Unlike the lesson plan for SL1/SL2, which provided a ‘bare-bones’ protocol for the science lab, no protocol is provided in the lesson plan for EL1/EL2. In other words, although both lessons are meant to be led in an open-ended fashion with multiple student-driven components, the science-based lesson plan did provide more structure to which it might be natural for teachers to easily tack on additional structure, to the point of hindering students’ opportunities for critical thinking (e.g., telling students what specific temperatures to test when measuring the influence of temperature, as a variable, on diffusion rate). In lieu of an experimental protocol, the engineering-focused EL1/EL2 lesson plan prompted teachers to set expectations for their students to reference and use what they learned from prior lessons to make informed design decisions. Indeed, most teachers in the focal qualitative group expected student groups to have their original designs approved by the instructor before they began building, and some of the teachers explicitly asked students to either write or describe their scientific and/or mathematical rationale for respective design decisions. It would seem that since no design protocol was provided in the INSPIRES lesson plan, that no teacher was in a particularly convenient place to provide a design protocol to students, therefore leaving it up to students to do their best in the open-ended exercise. Thus the nature of a well-written engineering curriculum, one that guides both teachers and students whom are engineering novices, provides a rich opportunity for multiple elements of pedagogical reform. This idea is reflected in the RTOP scores of the present study, whereby the INSPIRES engineering lessons exceed the science lessons; even teachers’ original design-based transfer lessons (T1, T2, T3a, T3b) largely exceeded the INSPIRES science lessons, even though reformed pedagogical strategies were written into the INSPIRES SL1/SL2 lesson plan.

Control Comparison

One strength of the present study and its research design is the inclusion of a control group of teachers that did not experience the INSPIRES PDs or enact the INSPIRES curriculum, but did participate in data collection at time points T1, T2, and T3a. Thus, comparisons could be made between the original, best-design or engineering lessons of teachers that received the INSPIRES intervention and those that did not. The assumption is that any given teacher, whether in the treatment or control group, may have had opportunities to pursue PD through the district or other means, which may influence their circumstantial growth in pedagogical reform or engineering practices. Therefore in order to begin teasing apart the role of the INSPIRES intervention from other interventions on such growth it is helpful to make comparisons between the teachers in the treatment and control groups. It is worth mentioning that although teachers in the control group participated in the study for three consecutive years, the overall duration was notably shorter such that video-recording data was only collected for T1, T2, and T3a. The timing of data collection of these three lessons equates to the timing of the treatment group counterparts. The delayed start in data collection from the control group is also why control comparisons were not included in the prior findings of the current research project, which focused on growth between Baseline, SL1, EL1, and T1 time points in the treatment group (Williams et al., 2019). Although T1 is the first instance of data collection for the control group, it is not exactly equivalent to the Baseline lessons of the treatment group because the timing of those lessons were not aligned. In other words, in the time spanning Baseline and T1 recording, teachers in both the control and the treatment groups could have received pedagogical training (outside of INSPIRES) that might influence their RTOP scores at time point T1. Another assumption of this study is that since all participating teachers may have the opportunity to growth pedagogically, through PD or other means outside of the INSPIRES intervention, it is expected that both control and treatment groups show some level of growth over three years. The extra enactment of an educative curriculum and training within INSPIRES PDs, however, is expected to yield greater growth or rate of growth in comparison to the control group. The results of the present study do not provide a clear cut distinction between control and treatment group pedagogical growth over time. Treatment RTOP scores for T1, T2 and T3a were each significantly higher than for treatment Baseline or control T1 scores. This suggests that the teachers in the treatment group demonstrated more pedagogical reform after receiving the INSPIRES interventions than at their starting point or that of the control group. Similarly, treatment group Baseline scores were not significantly different from control group T1 scores, which suggests that both groups of teachers utilized similar levels of reform at their respective starting points in the study, despite the chronological delay between the Baseline and T1 recordings. In other words, even though teachers in the control group theoretically had several more months of other district PD opportunities or other teaching-related experiences, as a group they still did not outperform the Baseline level of the treatment group on the RTOP instrument. Additionally, the T3a lessons of the treatment group significantly outperformed the T1 lessons of the control group. Together, the findings discussed above suggest that the teachers in the treatment group experienced more pedagogical growth than the control teachers, and therefore the combination of enacting an educative engineering curriculum and receiving ongoing pedagogical and content-specific support through PD result in pedagogical gains. Interestingly, however, significant differences are not detected between treatment and control groups within the same time points (i.e., at T1, T2, T3a, or T3b). While not significant, the treatment group did yield higher RTOP scores than the control group at each of these time points (Figure 2). It is speculated that the patterns of significant differences present and absent within this study may be influenced by small sample size (i.e., nine teachers in the control group) and the resulting statistical power, or the possibility that interventions must be sustained for three years or longer to begin prompting radical change. Of course, these lines of reasoning are not mutually exclusive and there may be other alternative factors at play that are less apparent at this time.

INSPIRES as an Educative Curriculum

The present study suggests that the INSPIRES educative curriculum could serve well as a tool to help teachers make reformed pedagogical changes. When INSPIRES lesson plans are followed closely by the enacting teachers, the educative components of the curriculum facilitates teachers' use of embedded pedagogical strategies. A high quality engineering lesson lends itself to the use of reformed pedagogy, as the content of an engineering design challenge typically demands critical thinking, problem solving, application, and reflection by the students. The INSPIRES curriculum is designed to place as much of the "burden" of grappling with content on the students, rather than the teacher. Every piece of the curriculum as written is deliberate. While some modifications and adaptations may be applied to accommodate the needs are different types of learners, such changes should be made with caution so as not to remove intended student-centered learning opportunities that promote engagement and deeper learning. For example, for students that require more structure to stay on task in an open-ended lesson, graphic organizers that lay out steps generally, without giving away specific answers, and that explicitly prompt students to write reflections, would arguably be a more reformed approach than giving students specific protocols, essentially turning SL1/SL2 into a confirmatory lab exercise. The INSPIRES curriculum was written with a broad array of student learners in mind. The INSPIRES lessons are meant to be a way to introduce all students to engineering principles and practices in an engaging and authentic manner. As a result, all students, regardless of their starting point, have an opportunity to grow in skills of applying math, science and technology principles to solve real world problems through design. Further, students may receive earlier exposure and proficiency in engineering skills to encourage their interest, pursuit, and success in college engineering tracks.

The INSPIRES educative curriculum can be evaluated here based on its relationship to the RTOP instrument. As mentioned earlier, lessons that are taught as intended (outlined by the lesson guides of the educative curriculum) yield high scores on the RTOP tool. The qualitative evidence described above, in combination with similar findings of prior work (Williams et al., 2019), suggest that the EL1/EL2 lesson plan was followed more closely that the SL1/SL2 lesson plan. The educative curriculum and teachers' experience using it is intended to be phenomena-first, as teachers demonstrate to themselves that they and their students are capable of benefitting from engineering content. Initially, by completing the INSPIRES lessons from a student perspective, teachers are actively engaged and accountable for their learning of the content and practices, until they reach a point where they have successfully designed, built, tested and redesigned a hemodialysis system from the materials supplied. Then, teachers use that valuable experience to inform their subsequent instruction of lessons to students. By working through an open-ended problem and coming out on the other side, teachers can feel the benefit of the content and the approach in an authentic way that would encourage them to provide that same experience to their students. Through these steps, teachers participating in the INSPIRES PD program realize the benefits of integrating engineering practices and reformed-based strategies into their existing biology or technology education classrooms.

Challenges and Limitations

Notably, some of the transfer lessons in this study yielded RTOP numerical values that suggested the encompassed pedagogies were 'less reformed' than expected. For example, while T1 resulted in an overall RTOP score significantly higher than the Baseline, it was higher than SL1 and lower than EL1, but neither significantly so (Tables 1 & 3). However, qualitative evidence helps to identify why teachers presented lessons that lacked elements of reform during T1, despite ongoing PD and having worked through the entire INSPIRES curriculum. The timing at which teachers were asked to present their "best design lesson" for T1 recordings coincided with the last month of the school year, and unfortunately, high stakes testing. To the researchers' surprise, many of the T1 lessons shared were structured as reviews or practice for students prior to their final exams or respective district high stakes testing. This was especially the case in the biology classrooms, as the subject matter is traditionally eligible for high stakes testing, while technology education content is not. While instances of documented teacher correspondence and T1 lesson plans suggest that some teachers opted for content reviews over inquiry-based lessons, the RTOP results reveal only a marginal difference between average biology and technology education teacher scores (Table 1). Additionally, none of the six teachers included in the qualitative review chose to share an assessment review as their T1 lesson.

Perhaps the more curious finding was the dip in RTOP score for T3b after a relatively high-scoring T3a (Table 1). Lesson T3b also did not score significantly higher than the Baseline lesson on the overall RTOP (Table 3), yet did on the individual subscales of Procedural Knowledge and Classroom Culture (Table 4). At first glance, it is surprising that the final opportunity for teachers to share how they can transfer learned pedagogies, at the conclusion of a 3-year study and intermittent PD experience, is not significantly higher than baseline levels when the preceding time point (T3a) is. Although, documented teacher correspondence and uncontrollable circumstances posed miscellaneous obstacles during the time frame when T3b lessons were recorded. For example, a misunderstanding regarding the scheduled time of data collection led to a missed opportunity. In this situation, a teacher had carefully prepared a lesson to share for T3b, but had enacted the lesson earlier in the day than the scheduled arrival of the data recorder, due to a last-minute adjustment in the school's A-day/B-day schedule. In another few cases, fire drills and inclement weather led to shortened class periods and last-minute lesson changes on the days of recording. Typically, these shortened or interrupted class periods led teachers to select more traditional, finite classroom activities in lieu of the time-intensive, open-ended design challenge that they might have preferred to share. Another possible explanation for the dip in relative reform at time point T3b, is the increased duration between the most recent INSPIRES PD and enactment of the final transfer lesson studied. Sustained PD may be important to the long-term efficacy of reformed pedagogy. While three years of PD can support significant change (Black et al., 2004; present study), more work is needed to discern the minimum length and frequency of PD required to maintain adopted change beyond the term of PD support.

While qualitative data provide some insight into why some transfer lessons were less reformed than anticipated by the researchers, a point can still be made regarding teachers' comfort with reformed pedagogies. That is, even in light of obstacles that prevented teachers from implementing their originally planned "best" engineering design lessons, in the examples described above there was the tendency to "fall back on" traditional lessons/pedagogy on any other day or circumstance. In other words, by working with the teachers through PDs, and corresponding with them at other times, there is qualitative evidence that suggests they were capable of producing very highly reformed lesson plans for their respective content classes. Yet, while teachers have succeeding in demonstrating growth in pedagogy according to data collected on a few pre-selected dates and lessons, it is worth exploring how commonly teachers implement their 'best pedagogies' or elements of engineering design, on their own accord and on typical days. While with planning and creativity it is possible to create design/inquiry-based/student-centered exam reviews or last-minute truncated lessons, perhaps more experience and more support through PD is required. If by the end of this three-year study, teachers were fully comfortable with implementing the reformed pedagogies highlighted in INSPIRES, and acknowledged the benefits of those pedagogies, then other day-to-day lessons could be quickly adapted (i.e., adding think-pair-share/probing questions/POEs, etc) to increase the student-centeredness. After all, RTOP scores are only meant to serve as a proxy for a given lesson's potential to serve as a rich learning opportunity for students.

Conclusions

Overall, the findings of this study suggest that the INSPIRES PD model is capable of effecting or producing teacher change in the area of reformed pedagogy. By exploring the INSPIRES lessons themselves, teachers may gain confidence in instructing new (engineering) content to their students. Those teachers that matriculated through the final year of the study (those that are the sample population of the present work) have experienced the positive impact that the INSPIRES curriculum has on students. Toward the end of the three-year study, the PD model shifted toward supporting teachers in areas of their own concern (such as how to prepare lessons for a newly adopted science curriculum by the district) and encouraged continued teacher investment in the pedagogical skill-training offered by both summer and academic-year PD sessions. Research has demonstrated that the longer the PD, the more effective it can be in changing teacher practices (Banilower et al. 2007; Boyle et al. 2005). After three years, transfer of skills is taking place within the teacher population discussed here, however, longer support may be needed to sustain pedagogical reform over time.

Recommendations

Next steps include aligning RTOP results with findings from other research instruments that measure teacher engineering content knowledge, student engineering knowledge, teacher attitudes toward teaching engineering, and infusion of engineering practices into a lesson. The broader scope of the research project has collected the aforementioned data and will work to piece together a bigger picture of the implications and affordances of the combined PD and enactment of the INSPIRES educative curriculum. In general, it is 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 teachers in implementing a strongly written engineering educative curriculum can allow the transfer of design-based pedagogy into teacher-developed curricula.

Acknowledgements

Jacqueline Krikorian, Jonathan Singer, and Christopher Rakes contributed equally to this work. 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, Rebecca Dowling and Ekaterina DiBenedetto. This project was funded by a Discovery Research K-12 National Science Foundation grant (DRL 1418183).

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 Researcher*, *25*, 6–8.

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the national science foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *44*(3), 375-395.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi delta kappan*, *86*(1), 8-21.

Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist*, *35*, 149-164.

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.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Resources*, *33*, 3-15. DOI: 10.3102/0013189X033008003

Boyle, B., Lamprianou, I., & Boyle, T. (2005). A longitudinal study of teacher change: What makes professional development effective? Report of the second year of the study. *School effectiveness and school improvement*, *16*(1), 1-27.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. T. Hallinan (Ed.), *Frontiers in Sociology in Education* (pp. 127-162). Dordrecht, Netherlands: Springer Netherlands.

Carr, R. L., Bennett IV, L. D., & Strobel, J. (2012). Engineering in the K-12 STEM standards of the 50 US states: An analysis of presence and extent. *Journal of Engineering Education*, *101*(3), 539-564.

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

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

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

Fishman, E. J., Borko, H., Osborne, J., Gomez, F., Rafanelli, S., Reigh, E., ... & 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.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of educational psychology*, *77*, 15-46.

International Technology & Engineering Educators Association. (2016). *Engineering by Design (EbD): A standards-based model program* [Website]. Retrieved from http://www.iteea.org/EbD/ebd.htm

Kelley, T. R., Brenner, D. C., & Pieper, J. T. (2010). Two approaches to engineering design: Observations in sTEm education. *Journal of sTEm Teacher Education*, *47*(2), 4.

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

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

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, CA: 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.

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.

Martin, W. G., & Gobstein, H. (2015). Generating a networked improvement community to improve secondary mathematics teacher preparation: Network leadership, organization, and operation. *Journal of Teacher Education*, 66, 482-493.

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*, 426–457.

Mehalik, M. M., Doppelt, Y., & Schuun, C. D. (2008). Middle-school science through design-based learning versus scripted inquiry: Better overall science concept learning and equity gap reduction. *Journal of engineering education*, *97*(1), 71-85.

Museum of Science. (2012). *EiE: Engineering & technology lessons for children* [Website]. Retrieved from http://legacy.mos.org/eie/

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

National Research Council. (2009). *Engineering in K-12 education: Understanding the status and improving the prospects*. National Academies Press.

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/

Noh, J., & Webb, M. (2015). Teacher learning of subject matter knowledge through an educative curriculum. *The Journal of Educational Research*, *108*(4), 292-305.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463-466.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American educational research journal*, *44*(4), 921-958.

Piburn, M., & Sawada, D. (2000). *Reformed teaching observation protocol (RTOP): Reference manual* (ACEPT Technical Report IN00-3). Tempe, AZ: 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.

Project Lead the Way. (2016). *Project Lead the Way: Preparing students for the global economy* [Website]. Retrieved from http://www.pltw.org

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning?. *Educational researcher*, *29*(1), 4-15.

Rakes, C. R., Bush, S. B., Mohr-Schroeder, M. J., Ronau, R. N., & Saderholm, J. (2017). Making teacher PD effective using the PrimeD framework. *New England Mathematics Journal*, 50 (1), 52-63.

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*, 331–350.

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. (2017). 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*, 43 (5), 789-818.

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*, 283–312.

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.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International journal of science education*, *28*(2-3), 235-260.

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

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, 102, 245-253.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference/William R. Shedish, Thomas D. Cook, Donald T. Campbell.

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

Timperley, H. (2011). *Realizing the power of professional learning*. McGraw-Hill Education (UK).

Williams, T., Singer, J., Krikorian, J., Rakes, C., & Ross, J. (2019). Measuring Pedagogy and the Integration of Engineering Design in STEM Classrooms. *Journal of Science Education and Technology*, *28*(3), 179-194.

Author Information

Tory H. Williams
University of Maryland Baltimore County
1000 Hilltop Dr. Baltimore, MD 21250
USA
Contact e-mail: *twilliams@umbc.edu*

Jacqueline Krikorian
University of Maryland Baltimore County
1000 Hilltop Dr. Baltimore, MD 21250
USA

Jonathan E. Singer
University of Maryland Baltimore County
1000 Hilltop Dr. Baltimore, MD 21250
USA

Christopher R. Rakes
University of Maryland Baltimore County
1000 Hilltop Dr. Baltimore, MD 21250
USA

Julia M. Ross
College of Engineering, Virginia Tech; 620
Drillfield Drive, Blacksburg, VA 24061
USA