Improving Middle School Science Achievement, Literacy and Motivation: A Longitudinal Study of a Teacher Professional Development Program

Gisele Ragusa Shaobo Huang Svetlana V. Levonisova
University of Southern California

Abstract

This research presents results of a longitudinal study of a three-year in-service teacher professional development program that was focused on improving grades sixth through eighth students' science achievement, science literacy and increasing science interest in five urban middle schools. The key program elements included: (1) a summer teacher academy, (2) academic year use of lesson study, (3) instruction using integrated science, (4) science notebooks use, (5) instruction on science literacy and effective use of informational texts, and (6) inquiry-based instruction. The study was part of a larger research effort that investigated the effects of a comprehensive middle school science and mathematics teacher professional development intervention on students' science and math achievement. The purpose of the research for this particular work is to examine the impact of teacher professional development on teachers' science instructional performance and science teaching efficacy, and in turn, its impact on their students' science achievement as well as their literacy in and motivation for science. The research utilized a cross-school comparison methodological approach, which examined middle school students' achievement trajectories and employed a quasi-control condition in the program's first year of implementation, and in remaining years – with multiple "doses" of an in-service teacher professional development intervention. Comparative and multi-level hierarchical linear modeling (HLM) was used for data analyses with multiple iterations of model fitting. Results of the research indicated that the professional development had multiple positive effects on the participating teachers and their students. School-wide teachers' science teaching efficacy increased for all participating teachers. The Academic Performance Indices (APIs) demonstrated growth for each of the five participating middle schools, and students' motivation for science and their science literacy had a positive effect on their science achievement. Furthermore, the participating students' achievement increased overall, with greater increases resulting from increased exposure to the teacher intervention (via proxy of the students' teachers). This teacher focused intervention is recommended for middle to large size middle schools and districts as it had moderate to highly positive effects on the teachers and students in the participating school district.

Keywords: longitudinal research, middle school, science achievement, science literacy, science motivation, science teaching, teacher professional development, teacher performance

Our nation's leadership has become increasingly concerned about its future workforces in science, technology, engineering and mathematics (STEM). Education in the STEM subjects from kindergarten through grade twelve (K-12) involves the inclusion of technology and engineering in mathematics and science school programs. Technology develops skills and abilities in adaptability, complex communication, non-routine problem solving, and systems thinking to "shape our material, intellectual, and cultural world" (Bybee, 2010, p. 31). These skills and abilities and how they are taught and learned within K-12 science classes are of primary concern to us. Therefore, we designed and implemented a teacher professional development intervention program to fully support our nation's future workforce with a focus on one of the most challenging age bands (grades 6-8) where STEM content dramatically increases in rigor.

The notion that teacher professional quality is a critical factor in student achievement has been widely acknowledged among researchers, policymakers, and teacher practitioners. Research has revealed that it can explain " . . . about 40 percent of the variance in students' learning and achievement – more than any other single factor, including student background . . ." (Rhoton & Stile, 2002, p.1). However, comprehensive inservice teacher education in STEM-focused pedagogy has been historically weak in United States public schools and by no means sufficient to contemporarily provide teachers with the necessary skills for success in their classrooms. Research on STEM education has indicated that having adequate subject matter knowledge, the "what" of teaching, is a necessary but insufficient condition to being an effective science teacher. Effective teachers also need knowledge and skills on "how" to teach the subject matter, and how to teach it in pedagogically inclusive ways so that all students, including those with increasingly diverse needs and abilities, will learn (Carlson, et al., 2019; van Driel, 2021). Finally, effective teachers need to understand students' development to be able to motivate and engage them and to help them reach their highest potential.

With the purpose to reinforce development of competent science teachers and to help retain them, the National Academies (2006) recommended improving science education by providing high quality professional development (PD) programs for science teachers. Science PD enhances improvement of the content knowledge and pedagogical skills that practicing teachers need to be effective in their classrooms. Accordingly, we have designed a compelling inservice middle school science teacher PD program aimed at improving teachers' subject matter and pedagogical practices using integrated science and discipline-based educational research practices (DBERs). For this work, we focused on the impact of a PD program on teachers' performance and efficacy in teaching science, and ultimately, on student achievement, in addition to their literacy in and motivation for science. Other aspects of our research have been addressed in additional publications.

The intent of our research was that by improving teachers' subject matter and pedagogical knowledge, their students' achievement would also improve. The project included a three-year intervention, encompassing teacher PD with integrated science. A data driven decision-making process was employed to develop the specific content of the teacher PD sessions and lesson study was utilized as an enabling structure for teacher collaboration, reflection, and self-study for improvement (Rozimella, 2020). Most importantly, the teacher PD had seven enabling components. The project represented a synergy of the components to specifically support the needs of the middle school teachers and their students in a historically low performing urban school district. The project was aligned to the State Department of Education's recommendations for middle school improvement: "Taking Center Stage-Act II: Ensuring Success and Closing the Achievement Gap for All of the State's Middle Grades Students. (CDE, 2021-rev.)"The project's enabling structures included: (1) university-based national research centers and laboratories; (2) a teacher training leadership team; (3) a content expert scientific and mathematics advisory team; (4) science teacher PD with a summer teacher academy and associated follow-up; (5) use of an adaptation of James Stigler's (2006) lesson study; (6) use of diagnostic teaching practices, science literacy, inquiry focused science, and data driven decision making; and (7) a focus on integrated science. The participating teachers received either two or three years of PD intervention, and therefore their 6th through 8th grade students received either a single, double or triple year dose of the intervention via their teachers depending on the year that they entered middle school. The specific interventions included within the PD consisted of: (a) use of academic language in science contexts, (b) instruction to improve science literacy and use of informational texts, (c) use of the scientific method in societally relevant middle school lab experimentation, (d) use of Cornell notes and science notebooks; (e) strategies for effective and efficient use of informational texts for improvement of content area literacy, (f) effective uses of technology in the classroom, (g) strategies for integrating science, (g) a teacher "boot camp" academy approach to improving teacher science content knowledge, and (h) strategies for nimble, data driven lesson design focused on inquiry and learning cycles. The teachers participated in annual weeklong summer teaching academies followed by grade level and subject specific mini-camps during the school year at school sites and in content groups. Lesson study was a major component of the teachers' PD during their academic years. A typical yearlong "dose" of the teacher intervention included 30 hours of a summer academy, followed by 15-20 hours per semester during the academic year (50-60 PD hours per year =1 PD dose).

Conceptual Framework

Review of research on teacher professional development in STEM not only addresses the significant challenges and needs of science teachers and their students, it has fully informed the design, development and impact testing of the described PD intervention. Therefore, the intervention addressed teacher content needs via boot camps with STEM content experts, included content-to-pedagogical disciplinary mentored practice, involved teacher self-study of their lessons (lesson study), and addressed students' science achievement, motivation for science, and content area literacy; all of which are addressed in the literature review that follows.

The Effects of Linking Content Knowledge to Science Teacher Professional Development

While there are numerous skills, strategies, and understandings one needs to teach science effectively, deep interconnected subject matter knowledge is critical to effective teaching in secondary schools. Researchers argue that inservice science teachers, particularly in middle schools, often fall short in their understanding of the contemporary science content that they are required to teach (Tretter et al., 2013). For example, in examination of the science content knowledge of 68 teachers in all grades from K-12 schools across Central Michigan, Parker, McConnell, and Eberhardt (2013) determined that while high school teachers were most likely to have deep science content knowledge, middle school teachers' science content knowledge was quite limited. To illustrate, only 13% of the studied high school teachers had low levels of science knowledge as opposed to 63% of the middle school teachers in the research. In contrast, 56% of the studied high school teachers showed high levels of science content knowledge in comparison to 23% of the middle school teachers.

Added to this, inservice science teachers often have sparse experience modifying curriculum content to include contemporary science related topics. Wenglinsky's (2000) study of the relationship between teacher PD and student achievement found that subject matter content knowledge is a necessary yet insufficient condition for high quality instruction. Wenglinsky's concluded that classroom practices, in the form of pedagogical content knowledge and teachers' use of higher order thinking skills, was a stronger predictor of student success than a teachers' subject matter content knowledge alone. As such, this research indicates that both content and pedagogical knowledge are critical to teacher and student success and thus provides a basis for our PD intervention's approach.

Shulman (1986) posits that it is a teacher's ability to recognize how to make content meaningful through disciplinary pedagogy that determines the teacher's skill to make transformative curriculum decisions for their students. Further, Mishra & Koehler (2006) conceptualized a framework for pedagogical content knowledge that described the complexities of the classroom as dependent upon the context of the teachers' everyday realities and the "…thoughtful interweaving of all three key sources of knowledge: technology, pedagogy, and content" (p. 1029). Effective teacher PD intended to innovate and change the quality of science curriculum must accentuate both subject matter content and disciplinary pedagogical knowledge, while emphasizing higher order thinking skills within the situational context of education (Burns et.al., 2018). This was at the core of our PD research design.

Teachers who fully understand the content area in which they teach, are more effective at producing higher achieving students. Disciplinary focused teacher PD can help to increase teachers' content knowledge by focusing on the specific subjects taught by the teacher, embedding teachers in environments with focused content, and assisting them in effectively teaching such content to their students. Through this type of PD, teachers can develop disciplinary pedagogical knowledge, in other words, an understanding of how students learn specific content. In Mundry's (2005) research, teachers participated in a PD experience where they explored science and mathematics case studies that integrated content and contexts for learning for teachers at the Far West Eisenhower Regional Consortium for Science and Mathematics. Mundry found that the students of the teachers who participated in the Eisenhower PD demonstrated significant gains in math and science test scores whereas the students of the teachers who did not participate had no significant gains on these math and science tests. Conclusively, research posits that PD that integrates content and pedagogy learning for teachers is highly likely to result in student achievement gains.

In related research, Sadler and colleagues (2013) investigated the relationships between teacher subject matter knowledge and student science gains. This study assessed both teacher and student content knowledge and it confirmed that teachers' science content knowledge is an important predictor of their students' content learning. Our study's design underscores this. Accordingly, we addressed practicing teachers' diversity in subject matter knowledge by embedding contextualized science content in each of the teacher professional development sessions so that the teachers would learn new science content that was contextualized for their middle school classrooms so that they could immediately apply their new content learning to their classroom environments through experimentation and investigation with their middle school students.

Need for Intensity and Multi-dimensional Structures in Teacher Professional Development

Professional development research has also indicated that particular intensity and structures are best suited for science PD. Importantly, inservice teacher PD needs to be sustained and intensive for it to translate into student achievement gains (Wei, et al., 2009; Darling-Hammond, 2012, Ambussaidi, et al., 2019). Through a meta-analysis of PD interventions, Wei and colleagues (2009) found that programs offering 30-100 contact hours over six to twelve months had a significant impact on student achievement gains. For programs offering an average of 49 hours in one calendar year, student achievement rose by 21 percentile points (Wei et al., 2009). In one specific reviewed study, science teachers participated in a 100-hour summer PD where they explored a scientific phenomenon, developed a theory to explain the phenomenon, and applied such theory and content to new contexts. The teachers then developed units of study around this phenomenon and taught them to their peers, thereby supporting the notion of to teach is to learn. When the students of these teachers were tested, they scored 44% higher on achievement tests on average than the students whose teachers did not participate in the PD (Wei et al., 2009). This research review indicates that sustained and intensive PD can have a significant positive effect on student academic outcomes. It is for this reason that our study's intervention included both a summer immersion component of 60 + hours and ongoing fall and spring semester follow-up (30-50 hours; Granger et.al., 2018).

Lesson Study and Its Link to Teaching Effectiveness

Particular types of teacher PD structures have yielded more impactful results. Several studies suggest that teaching effectiveness is improved through a "lesson study" approach to teacher PD (Akerson et.al., 2017; Cajkler & Wood, 2016). Lesson study, according to Stigler & Hiebert, (2016; Stigler, 2006), refers to a PD process whereby teachers closely examine their lessons with a reflective focus on addressing student need via data-driven decision making, creating powerful and relevant curricula, and reformed lesson design. Lesson study goes beyond collaboration to co-planning and observing actual lessons of peer teachers with a focus on student thinking and learning (Kohlmeier, et.al., 2020). In the lesson study model, teachers learn together. They plan, observe, and refine their lessons designed to make real their long-term goals for their students' learning and development. A key component of lesson study is observing and teaching lessons, which are improved collaboratively (Jones & Gallen, 2016; Rozimella, 2020). This compels teachers to examine their own practice in depth in the context of student learning, connects them with their students and their professional community, and inspires them to continuously improve, (Kanelloupou & Darra, 2018a, 2018b). This model of teacher PD has been applied widely and successfully in Japan and has more recently been initiated by teachers at many sites across the U.S. and beyond (Saito & Atencio, 2013; Schipper, et.al. 2017). It is especially applicable to science and mathematics education (McNally, 2015, Kohlmeier et. al., 2020, Kayapinar, 2016).

Gerard, Varna, Corliss and Linn (2011) found that PD demonstrates greater teacher instructional improvement when the program uses inquiry investigations. Inquiry investigations in this context consist of comparing different curriculum and pedagogical techniques, improving lesson plans, discussing student ideas in a specific subject area, and connecting student ideas to instruction. This research parallels a "lesson study" approach to PD, which has been found to be effective in improving both teacher practice and student achievement, (Stigler, 2006). The approach is based on a knowledge integration framework that focuses on building upon learner ideas by utilizing evidence to add new content. In a meta-analysis of forty-three PD intervention studies in science education, teachers in grades 6-12 were compared on the impact of PD on their teaching, based on the knowledge integration framework. The analysis revealed that more than 68% of teachers in the PD programs enhanced students' inquiry science learning experiences when the PD interventions enabled participating teachers to follow the knowledge integration framework and was sustained for more than 1 year (Gerard et.al, 2011). Using inquiry investigations, both for teachers and their students, can therefore, have a strong impact on teachers' instructional practices.

Lewis, Perry, and Hurd (2009) applied the lesson study PD approach during a session of a 2-week summer workshop for teachers in a large urban school district. In this study, teachers were involved in a focal lesson study group. In the first phase of the lesson study, the teachers studied their state's content standards, and discussed and solved problems. In subsequent phases, the teachers selected, observed, and collected student data from a research lesson, and then discussed, revised, and re-taught their lesson to another group of students. Teachers were videotaped as they wrote a lesson plan, taught and observed the research lesson, revised the lesson plan, and re-taught the lesson. Data were also collected from group meetings, student work, field-notes, and follow-up conversations. Results of this multi-dimensional study indicate that teachers' disciplinary instruction was improved through lesson study by changing not only teachers' content knowledge, but also their pedagogical practices as well as their community, and teaching-learning resources.

Listyani, Widjajanti, and Susanti (2008) also found numerous positive impacts of lesson study on teachers' competence. Teacher competence was increased, which included pedagogical, professional, and social competence. In this particular research, the lesson study activities included planning, implementation, observation, and reflection. The teachers discussed measurement tools for student learning, lesson plans, students' work artifacts, and evaluation instruments in their lesson study. Each teacher acted as a teaching model for two rounds of lesson study and observed other members' teaching processes for eight rounds through the lesson study processes. Data were collected from student questionnaires and interviews, teacher interviews, and in-class observations at the end of the lesson study. Results of this research demonstrated that greater than 80% of the students were more involved in their learning activities after the teachers completed lesson study. The teacher participants had better classroom management and improved lesson design skills resulting from the lesson study experiences as well. It is from these combined reviewed results that our PD program incorporated lesson study as a critical component of our study's teacher intervention.

The Impact of Science Literacy Instruction on Student Learning

Not only are the content and structures important considerations when designing teacher interventions; the particular instructional approaches taught during the PD are also critical. Literacy is a particularly challenging area for students in U.S. schools. Content area literacy and especially science literacy using informational texts is of pronounced difficulty for students, especially in middle schools (Mahan, 2020). Therefore, in addition to teaching content knowledge to teachers, guiding teachers on how to teach literacy with science informational texts was an essential component of our teacher PD. Empirical research highlights the benefits of integrating literacy instruction into inquiry-based science in secondary contexts (Zucker, Noyce & McCullough, 2020). To illustrate, in a quasi-experimental study, Fang and Wei (2010) investigated the effects of an inquiry-based science curriculum that integrated explicit reading strategy instruction and use of high quality science trade books on the development of science literacy among middle school students. The intervention condition for this particular study included two components of reading called Inquiry-based Science Plus Reading: explicit reading strategy instruction, and use of scientifically focused "trade books" with a reading response sheet with guided teacher discussion about the books. Several measures of the impact from this intervention on the students' science literacy were used: Norris and Phillips' (2003) new conception of science literacy (measured students' science literacy development in both the fundamental and the derived senses), the Gates-MacGinitie Reading Tests, curriculum-referenced science test (assessed students' derived sense of science literacy), and the students' academic year science grade. The results of the study indicate that the Inquiry-Based Science Plus Reading (intervention) group significantly outperformed the control group in fundamental components of science literacy including science vocabulary and informational text reading comprehension. The results demonstrated that an inquiry-based science curriculum that infuses explicit reading strategy instruction was more effective than an inquiry-based science only curriculum in developing students' science literacy.

In comparatively similar literacy research, Anthony, Tippett, and Yore (2010), investigated the benefits of embedding explicit literacy instruction into middle school science. This research utilized "a community-based, opportunistic, engineering research and development approach to identify problems and concerns and to design instructional solutions for teaching middle school science." Accordingly, diagnostic instruction in science literacy was used as an intervention. Results of this research suggest improved performance associated with science reading and improved science literacy strategies. As such, the decision for including deliberate content area literacy components into our teacher PD intervention was noteworthy. This research underscores both the depth and complexity of our teacher professional development research that is described in the pages that follow.

Method

Data Collection

Study Setting

This study took place in an urban mid-sized public school district with five middle schools. Three of the middle schools are traditional middle schools, thereby serving students in grades 6-8, one school site is a K-8 school, and one school is a 4th-8th grade school. Only sixth through eighth grade teachers and classrooms in the K-8 and 4-8 schools were a part of our study's intervention. The school district is situated amidst a community that is struggling significantly both economically and resource-wise, and therefore the educational and materials needs in the district are great. All schools that were targeted for the intervention, were in federal Program Improvement (PI) status at the onset of the intervention and accordingly, their academic performance indices (API) were low. In this school district, the students' academic needs were great at the start of this intervention, with 55-82% of 6th-8th grade students scoring non-proficient (basic to far below basic) on their state standardized achievement tests.

Study Sample

The study sample included two groups: middle school teachers, and their sixth through eighth grade students. There was a total of 64 teachers in the study sample including general education science teachers and a handful of special education teachers. Principals and other site and district administrators also participated in the intervention; however they were not a part of the study sample. The student sample consisted of 5,505 students. The participating students were primarily of Latinx/Hispanic decent, however there was an ethnic mix in the sample. The distribution was 88.3% Latinx/Hispanic, 7.6 % African American, 1.4% White, and the remaining 2.7% from other ethnicities. Linguistically, 40.1% of the students in the sample spoke Spanish as a primary language at home and 33.2% of the students were categorized as limited English proficient at the start of the intervention. Additionally, 84.3% of the students received free or reduced lunch across the five middle schools. The students' academic needs were pronounced prior to the start of the intervention, with 59-68% of all 6th-8th grade students scoring non-proficient (basic to far below basic) on the state standard tests in English language arts (ELA), 57-75% 6th-8th grade students scoring non-proficient (basic to far below basic) on the state standard tests in mathematics, and 55-82% 6th-8th grade students scoring non-proficient (basic to far below basic) on the state standard tests in science. This achievement gap persisted for six years prior to the intervention and was particularly dismal for ethnic minority student groups (~13% lower than non-minorities) and English learners (~10-23% lower than "English only" students).

Recruitment Process

In terms of recruitment for the intervention, both the teacher and student samples were recruited through the school district administration and therefore recruitment was inclusive. All 6th-8th-grade science teachers in the district and their students participated in the intervention. Because the data provided by the school district were masked for identification for human subject protection (and this decision was made a priori to the intervention), a parent-child "opt out" procedure was employed in the recruitment design. No families chose to opt out of the intervention. Teachers were compensated for participation in the intervention via the project funds directly during the summer teacher academy and via releases with substitute teachers during the academic year, as academic year intervention sessions were conducted during the teachers' contracted day.

Intervention and Differences in Control/Comparison Group Condition

The intervention was targeted at middle school teachers in the district with an intent to positively impact their students' achievement in science. Therefore, the teacher PD had seven enabling components that were well established in the supporting university and endorsed by the school district. The program represented a synergy of the components to specifically support the needs of the middle school teachers and their students. The project was aligned to the State Department of Education's recommendations for Middle School improvement: "Taking Center Stage-Act II: Ensuring Success and Closing the Achievement Gap for All of the State's Middle Grades Students" (CDE, 2021-rev). The enabling structures included: (1) university's national research centers and laboratories; (2) a teacher training leadership team; (3) a content expert scientific advisory team; (4) science teacher PD using a summer teacher academy and associated follow-up; (5) use of a modified form of Stigler's lesson study; (6) use of diagnostic teaching, science literacy, inquiry focused science, and data driven decision-making; and (7) a focus on integrated science. The teachers received either two or three years of PD intervention, and therefore their 6th through 8th grade students received either a single, double or triple year-long intervention dosage depending on the year that they entered middle school. The specific interventions consisted of: use of academic language in science contexts, science literacy, real-life problem solving, use of the scientific method in societally relevant middle school lab experimentation, use of Cornell notes and science notebooks, strategies for effective and efficient use of informational texts, strategies for improvement of content area literacy, effective uses of technology in the classroom, strategies for integrating science, a teacher "boot camp" approach to improving teachers' content knowledge (with content experts), and strategies for nimble, data driven lesson design focused on inquiry and learning cycles pedagogical structures. The teachers had weeklong summer teaching academies followed by grade level and subject specific mini camps during the school year at school sites and in content groups. Lesson study was a major component of the PD. A typical yearlong "dosage" of PD included 30 hours of summer academy, followed by 15-20 hours per semester during the academic year (50-60 PD hours per year = 1 dose).

The research team included a one-year quasi control condition in the research design. To facilitate this, the three traditional 6th-8th grade middle schools began implementation in Year 1, and the two non-traditional schools (gr. K-8 and gr. 4-8) served as quasi "control" schools during that year, thereby using a "business as usual" educational condition, and began PD intervention the following year (Year 2). Once the implementation of the full intervention across all five schools began, the intervention was delivered in groups and therefore no differences in treatment were noted (except for dose because of the deliberate delayed start of implementation for two of the schools). There was no randomization of treatment in the study. The selection of schools and years of implementation was made at the request of the school district administration because the district was transitioning from junior high school models to middle school models prior to the start of the intervention.

Research Questions and Associated Hypotheses

This study responds to four important research questions and associated hypotheses:

- What is the impact of lesson study focused, content rich science inservice teacher PD on middle school teachers' science teaching efficacy?
  - Hypothesis: The described teacher PD will increase participating teachers' science teaching efficacy.
- What is the impact of the described teacher PD on the middle school science teachers' students' science achievement?
  - Hypothesis: The described PD will increase the participating teachers' students' science achievement.
- What is the impact of the described teacher PD on middle school students' science literacy?
  - Hypothesis: The described PD will increase students' science literacy.
- What is the impact of the described teacher PD on middle school students' science interest and motivation?
  - Hypothesis: The described teacher PD will improve students' interest in and motivation for science.

Outcomes Measurement and Instrumentation

The outcomes for this research were measured both at the teacher and student levels. Primarily, instrumentation was standardized, statistically reliable, and highly validated using item response theory (Wilson, 2011). Because the research was a mixed design study, field notes from lesson study group meetings, rubrics and planning forums, were also used and needs assessment (open and close set) questionnaire data was collected to add depth to the breadth of the study's data.

Instrumentation For Teachers

Instrumentation for teachers included:

- Science Teaching Efficacy Beliefs Instrument (STEBI-R): This is a Likert-type questionnaire that measures teachers' science teaching efficacy.
- Lesson Study Scoring Rubric: This is a multi-dimensional observational rubric (scaled through full implementation of lesson study), called the Teacher Performance Observational Rubric (TPOR; adapted from PACT; 8 points possible -across rubric dimensions).
- Interview Protocol: These were periodic focus group interviews that were conducted with teachers in the intervention to assign voice to their experiences during the intervention.
- Lesson Study Structured Field Notes: These were notes that were taken of teacher interaction and discussion during the lesson study planning and implementation process.
- Teacher Feedback and Ongoing Needs Assessment Questionnaire: This was an electronically administered questionnaire that provided formative feedback and needs identification from teachers to the Teacher Leadership Team (TLT).

Instrumentation For Students

Instrumentation for the students contained:

- State Standards Tests (SST) in Science (Gr 8): These are the state adopted standardized achievement assessments.
- District Benchmark Exams in Science (Gr. 6-8): These are the target school district's science benchmark examinations for grades 6-8 that are administered quarterly. They are criterion referenced with district calibrated, accepted "cut scores."
- Qualitative Science Literacy Inventory: This is a science literacy measure that was designed, tested and validated using IRT (Wilson, 2011) by one of the authors that was administered as a pre-post comparison annually. The teachers administered this inventory at the start and end of each academic year throughout the intervention period.
- Motivation for Science Questionnaire: These Likert-type questionnaires were designed, tested, and validated using IRT (Wilson, 2011) by one of the authors and administered by the research team via the teachers as an annual pre-post at the start and end of every academic year of the project. This questionnaire contains 10 subscales associated with motivation, curiosity, engagement and efficacy in science. Students received the science motivation questionnaire during their science class.

Instrument Reliability and Validity

The study's instrumentation has been tested for reliability and validity. The TPOR has been used as a primary instrument for one of the author's federally funded research. It consists of six teacher instructional performance dimensions including planning, assessment, instruction, reflection, academic language, accommodations for diverse learners (each with one or more sub-dimension,) and its reliability is robust (Cronbach's alpha = .93; Ragusa, 2011). It was modeled after a combination of the Performance Assessment of State's Teachers and Ball and Forzani's (2010) teacher observational assessment. It was tested for validity and reliability using Wilson's four building blocks of item response theory (Wilson, 2011). The students' science literacy inventories (grades 6-8) have also been tested for validity and reliability using Wilson's item response theory. They have strong reliability as well (Cronbach's alpha = .82-.86; Ragusa, 2011). The STEBI-R has established validity and reliability through studies conducted by the instrument's creators, Riggs & Enochs, (1990). Reliability of the instrumentation was assessed using the Kuder-Richardson 20 formula, which is appropriate for examining internal consistency for binary (i.e., "right" versus "wrong") responses (Fraenkel & Wallen, 1990). Collectively, in terms of statistical power of the full set of instruments, given that there were 5,505 in the student sample, statistical power was robustly achieved (Cohen, 1992).

Analytical Methods

The study employed a mixed methods research design. Accordingly, data analyses were both quantitative and qualitative. In addition to conducting descriptive analyses to illustrate the diversity in the study's sample, pre-post comparisons, including t-tests with effects (Cohen's *d*) were computed. Correlation analyses were conducted as a precursor to multivariate and multilevel hierarchical linear modeling (HLM). Multiple means of model fitting were used in the HLM analyses (see results for particulars and model development) and SPSS (version 22) and HLM (version 7.1) software were used for these analyses.

For the qualitative portions of the research, data categorization of teacher data with frequency distributions was conducted. Specifically, open-ended responses to the teacher questionnaires (needs assessment and evaluative feedback from PD), lesson study and planning observations, and field notes were analyzed using well established thematically focused qualitative analyses and NVivo (version 10) was used for these analyses. The data were coded and thematically categorized using a constant, comparative method, (Lincoln & Guba, 1985). Special attention was paid to disconfirming evidence and outliers in data coding, as well as elements of frequency, extensiveness, and intensity within the data. Ideas or phenomena were initially identified and flagged to generate a listing of internally consistent, discrete categories, followed by fracturing and reassembling (axial coding) of categories by making connections between categories and subcategories to reflect emerging themes and patterns in the data. Categories were then integrated to form grounded theory and aligned with existing teacher development theory using selective categorization to clarify concepts and to allow for response interpretations, and conclusions associated with the teachers' perceptions of the successes and challenges of the lesson study, the PD and their students' changes, strengths and difficulties. Frequency distribution of the coded and categorized data was computed. The intent of this intensive qualitative analysis was to identify patterns, make comparisons, and contrast one person or groups' discussion, action, and voice with another throughout the project to provide a complete picture of the intervention qualities.

Limitations of Data and Analytical Methods

The data for this research was limited by several factors. First, all student level data was collected by the participating classroom teachers. Therefore there was some missing data that was accounted for by reliable and widely accepted statistical procedures for managing such data. In terms of the analytical approach, because multivariate approaches and multi-level hierarchical linear modeling were used, limitations were not profound. The primary limit analytically emerged when, in an attempt to build and fit a three-level model (students nested in teachers' classrooms and nested in schools), it was determined that due to the shared and homogenous nature of the five schools' characteristics, no significant effects were noted statistically at Level III of the model, thereby causing a need to return to a two-level hierarchical model. The interpretation of this analytical structure is described in the results.

Results

Indicators of Successful/Unsuccessful Study Process

Indicators of the success of this study included low attrition rates in students and no attrition in teachers. Additionally, while the teacher sample size was somewhat modest (N = 64; however not so modest given the context and scope of the study and the nature of public middle schools), the student sample size was quite robust (N = 5,505). The intervention had a very low dropout rate for students; however the research team did have to eliminate some student data from the sample due to missing data and data matching difficulties associated with the relative anonymity of the data (the team employed a unique identifier formula for the data in which the students were responsible developing the identifier based on their first name, last name initial, mothers' month of birth, teacher, subject and period of class).

Descriptive Data Describing Index of Implementation (Fidelity)

The project leadership team achieved maximum fidelity of implementation of the intervention because at

Grade 6	Grade 7	Grade 8
Mean = 6.21	Mean = 5.93	Mean = 6.04
SD = 0.88	SD = 1.26	SD = 1.29

Teachers were scored only one time during the intervention and therefore there was no pre-post intervention comparison.

Table 1. TPOR Scores by Grade (8-Pt. Rubric Score; Between Years 3 & 4)

Teaching Enthusiasm	49 (44.1)	63 (36.2)	" I was beginning to feel very burnt out." "Now I'm feeling invigorated and raring to go in my teaching. I look forward to the days."
Pedagogical Language	17 (15.3)	42 (24.1)	"I find using ACT work best." "I like using the 5 Es as it allows for exploration and student inquiry."
Collaboration	22 (19.8)	27 (15.5)	"Lesson study gives us time to work across schools. We never had this before." "My favorite PD is working together!"
Help Seeking	19 (17.1)	13 (7.5)	"I am glad we are able to keep (Name, math coach). I rely on her daily!" I feel supported and know I can get help when I need it now."
Help Provision	4 (3.6)	29 (16.7)	"I don't care if you have a look at my video. I'll bring it by along with some materials and we can plan the next couple of weeks."
TOTAL	**111 (100)**	**174 (100)**	

Table 2. Teacher Perceptions of Intervention Across Time

the teacher level, the project director was fully responsible with her team for the design and implementation of the teacher PD. This was implemented in consultation with the teacher leadership team, which remained in place throughout the planning and implementation period of the intervention.

Estimates of the Intervention's Effect on all Outcomes (with Subgroups)

Given that this research involved teachers and their impact on students, we wanted to measure the impact of the teachers' intervention and change in instructional practice and knowledge on their students' achievement, motivation and interest. Accordingly, the results are divided by study population, (below), and then the combined/interactive effects are described and illustrated. Within the teacher related results, both quantitative and qualitative results are presented.

Teacher Effects

Teacher effects resulting from the intervention relate both to teaching efficacy (measured by the STEBI-R) and teacher performance (measured by an observational rubric, called the TPOR). These results are interesting and diverse. All teachers participated in lesson study and the lesson study consisted of preparation of collaborative lesson plans, assessment plans, videotaped teaching events, debriefing, scoring sessions and plan revisions. The TPOR was scored multi-dimensionally on an 8-point rubric with a score of eight being the highest possible score. The mean TPOR scores by subgroup of science teachers are the following:

Teachers were scored via lesson study only one time per teacher during the intervention and therefore there was no pre-post intervention comparison.

With regard to teachers' efficacy for teaching science, there were moderate differences between the pre and post intervention scores: *Grade 6 Teachers:* $M_{pre} = 3.03$, $SD = .231$, $M_{post} = 3.62$, $SD = .418$ $t(22) = 6.99$; Cohen's $d = 1.723$; *Grade 7 Teachers:* $M_{pre} = 3.17$, SD = .224, $M_{post} = 3.82$, SD =.222 $t(29) = 10.3$; Cohen's $d = 2.355$; *Grade 8 Teachers:* $M_{pre} = 3.19$, SD = .246, $M_{post} = 3.75$, $SD = .380$ $t(32) = 10.7$; Cohen's $d = 1.751$. There were also some noteworthy variations in results across teacher subgroups. Specifically, teachers with less than 2 years teaching experience significantly improved (mean difference = .49, $t(5) = -3.726$; $p < .05$) during the intervention. Teachers who taught only 7th grade improved at a somewhat lower yet still at a statistically significant rate (mean difference = .30, $t(6) = -5.81$; $p < .001$). Teachers who taught only one grade had a significant increase in teaching efficacy (mean difference = .20, $t(15) = -3.704$; $p < .01$). Additionally, when correlated with the Teacher Performance Observational Rubric (TPOR), with 8 points possible, the teachers' instructional performance results were highly correlated with their teaching efficacy post intervention ($M = 5.47$, $SD = 11.03$; $r = .47$).

Qualitatively, the teachers' perception of the PD, their reactions and perceived progress were quite interesting and remarkable. These results are illustrated below in Table 2 including samples of teacher quotes taken from field note discussions.

Teaching Enthusiasm. Teaching enthusiasm is an element of teacher motivation that is reflected by the subjective value teachers place on teaching. It is often reflected through verbal and nonverbal expressiveness that influences the engagement in a learning environment as well as teaching-related enjoyment perceived by teachers and their students (Lazarides, Fauth, Gaspard, & Göllner, 2011; Keller, Hoy, Geotz, & Frenzel, 2016).

Pedagogical language refers to the discussion of instructional practices that teachers often engage in with one another while they reflect on their teaching practices and especially when they are supporting one another or being supported during professional development opportunities.

Collaboration. Collaboration applied to teaching refers to a reciprocal learning atmosphere for teachers to improve teaching practices over time. It encourages sharing professional knowledge and experiences and often results in higher teacher motivation, mental health, and job satisfaction, as well as significant improvement in student learning and achievement (Kolleck, 2019; Ostovar-Nameghi & Sheikhahmadi, 2016; Vangrieken, Dochy, Raes, & Kyndt, 2015).

Table 2 (to the left) provides a description and frequency distribution of the teachers' comments and discussions in which they engaged during the lesson study sessions and other professional development opportunities. These comments were qualitatively analyzed to "quantify" the rich and reflective discussions that the teachers engage in during their learning processes as they were supported by our research team and by one another. To illustrate the content and depth of the discussions, example excerpts of the thematically analyzed teacher discussions are provided in the table.

The combined teaching efficacy and qualitative results illustrated above indicate that both teacher attitude toward teaching and teaching efficacy; in other words their confidence specific to teaching math and/or science, increased over time. Additionally, the teachers became more engaged and reported being more knowledgeable about their teaching practices. Examples of such engagement is indicated in Table 2.

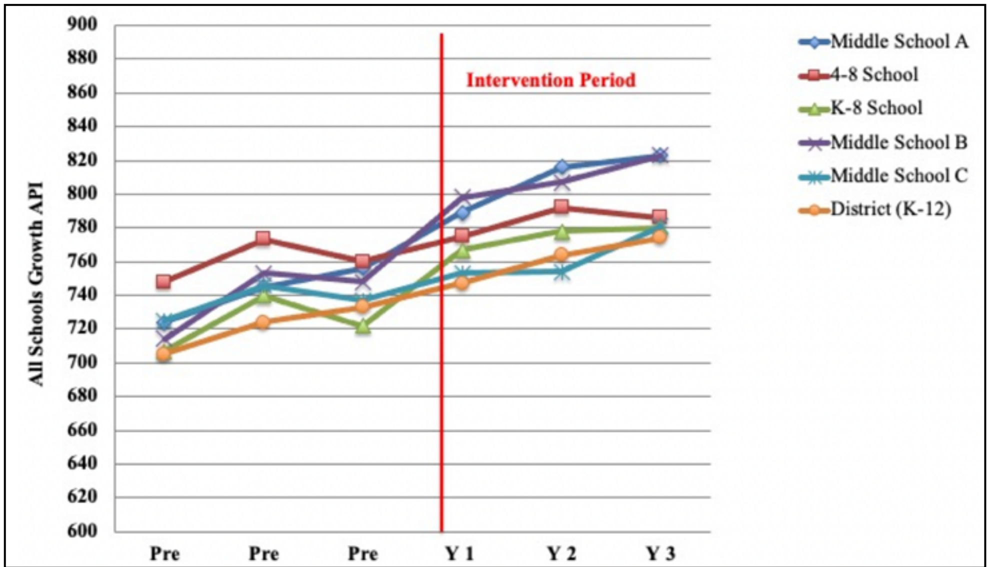


Figure 1. Cross School Comparisons – Academic Performance Index (API) Pre and During Intervention

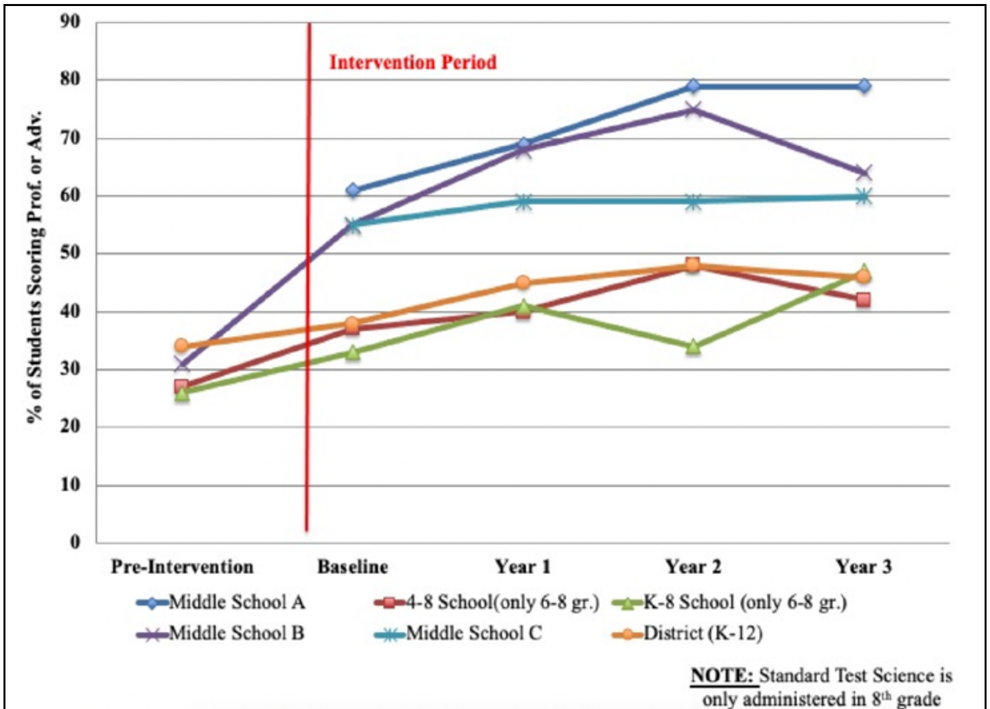


Figure 2. Student Achievement Standard Scores Science–Cross School Comparisons

Student Effects

The student results were also comprehensive, diverse and impactful. With regard to changes in student achievement, while the results of the students' State Standards Tests (SST) in science were variable, on the whole, all five intervention schools had gains between the pre intervention period and final year of the intervention. The same was true when tracking the trajectories of the schools' Academic Performance Index (API.)

As illustrated by Figures 1 and 2 (above), the 8th grade students in the sample experienced steady gains in science test scores across the three intervention years. The test score growth varied across years. In the baseline year, two of five schools were below district and state percentage scoring proficient or advanced in science. Between the first two intervention years, in science, the intervention schools had 8.33% growth in percentage of students scoring proficient or advanced while quasi control schools had a 5.05% increase in percentage of students scoring at proficient or advanced on the SST. Between the first and last years of intervention, while there was variability in growth, the percent of growth in students scoring proficient or advanced on standardized science tests was 9.96% and, the percentage growth of students scoring proficient or advanced was 12.58%, both of which were above the district growth percentage and the state growth percentage.

Science Literacy

With regard to science literacy, the students' sample had variability in changes across time in the study. These results are presented below in Table 3.

These results indicate moderate variability across years. Comparisons across grade levels in the sample were not made because the students' science literacy inventory content varied year-to-year as the three dimensions of the exam got successively more difficult with each grade level and because the content of the exam in each grade was aligned with the particular science for the grade level (e.g. the 7th grade science literacy inventory is aligned with life science content).

Motivation for Science

With regard to science motivation and associated sub-constructs for this measure, results indicate that there were increases in intrinsic motivation across years for the students, with some variability across sub-constructs. Specifically, over time, there were larger gains in intrinsic motivation and decreases in extrinsic motivation for the participating students. In other words, the students in the study sample improved their ability to be motivated to engage in science without external rewards of any type, which is supported as a positive effect in the literature (Guthrie, McRae & Klauda, 2007). For example, Science Efficacy $t(2315) = 3.07$; $p < .001$; Cohen's $d = .47$; Science Challenge $t(2315)$

Grade 6 Year 1 (1 Dosage): M_{Pre}= .5044, SD =.1855 M_{Post}= .5461, SD =.1882 $t(2897)$=10.85, p<.001, Cohen's d = 0.2245	**Grade 8 Year 1 (1 Dosage):** M_{pre}= .6995 SD=.1636 M_{post}= .5945 SD=.1980 $t(974)$=14.37, p<.001, Cohen's d = 0.5781
Grade 7 Year 1 (1 Dosage): M_{pre}=.6483, SD=.1676 M_{post}=.5870, SD=.1920 $t(1029)$=9.202, p<.001, Cohen's d = 0.3401	**Grade 8 Year 2 (2 Dosages):** M_{pre}= .6055, SD =.1798 M_{post}= .5496 SD =.2429 $t(915)$=6.583, p<.001, Cohen's d = 0.2615
Grade 7 Year 2 (2 Dosages): M_{pre}= .5857, SD =.1753 M_{post}= .5844, SD =.1836 $t(1766)$=.295, p>.05. Cohen's d = 0.0072	**Grade 8 Year 3 (3 Dosages):** M_{pre}= .6028, SD =.1691 M_{post}= .6071, SD =.1305 $t(763)$=.7630, p >.05, Cohen's d = 0.0284

Table 3. Science Literacy Across Grades and Years

= 2.64; $p < .01$; Cohen's $d = .29$; Science Curiosity $t(2315) = 2.41$; $p < .05$; Cohen's $d = .21$.

Linear regression analyses of science motivation, using the standardized science test scores as a dependent variable, were conducted as an analysis. These results indicate the following: Science Motivation predicted student achievement in science in the sample (Grade 8 science SST; 6th and 7th grade students do not take science SST).

- $R^2 = 0.14$, $F(10, 3021) = 57.50$, $p < .001$

This result is supported in the literature by motivation/achievement research conducted by Guthrie and colleagues (2007).

Combined/Integrative Effects

After completing analyses for both the teachers and students in the research sample, and because the primary goal of the project was to test the impact of a teacher professional development intervention on student achievement, hierarchical linear modeling was used for analyses. The research team proceeded through various iterations of model fitting following well-established research by HLM founders Stephen Raudenbush and Anthony Bryk. As such, initially, a three-level model with students' nested in teachers' classrooms, which were nested in school sites was utilized. After testing this model (and its null counterpart) it was determined that there were no statistically significant school effects on the first two levels of the model *(interceopt1/intercept2: Chi Sq. = .01726, p >500)* and therefore, the analysis shifted to a two-level hierarchical linear model with students nested in their teachers' classrooms. With this approach, the models in statistical notation and their associated complete results are illustrated below in Tables 4 through 7. Importantly, the model was run with outcome (dependent) variables for science benchmark exams (district level) for grades 6-8 and science standardized test scores for grade 8 (grades 6 and 7 do not take the statewide standardized science test). All variables for each model are labeled within each model. Both unconditional and conditional models are illustrated below with model equations in numbered order. Dummy coding was used for variables including teacher credentials (whether the participating teacher had a single subject credential or not) gender and ethnicity. Both aggregate and non-aggregates of predictor variables were utilized in the models. Group centering was employed and is clearly delineated in each model. Students' English Language Arts (ELA) standardized test scores were loaded into the science HLM models as we hypothesized that this variable might impact student achievement in science. This was confirmed via a priori correlational analyses between the variables.

Model	*Fixed Effects*	*Coefficient*	*t-ratio*	*P*	*Reliability*
Grade 6 Model	Intercept 1, β_0				
	Intercept 1, γ_{00}	0.588	36.999	.000	0.963
Grade 7 Model	Intercept 1, β_0				
	Intercept 1, γ_{00}	0.633	60.915	.000	0.918
Grade 8 Model	Intercept 1, β_0				
	Intercept 1, γ_{00}	0.692	40.141	.000	0.951

Table 4. Unconditional Model Using District Science Benchmark Exams (SUM mean) as Dependent Variable

HLM Models - Science Benchmarks (SBK) as Dependent Variable

Full Unconditional Model

Student - Level Model:

$$\textbf{Science Benchmark}_{ij} = \beta_{0j} + r_{ij} \quad (1)$$

Teacher - Level Model:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

Results of the hierarchical linear modeling indicate some interesting effects. First, using the participating district's science benchmarks scores (a measure of achievement) as an outcome (dependent) variable, for 6th grade students, teaching efficacy had a positive effect on student achievement (*03 = .092, t(2.225); p < .05*). Additionally, students' intrinsic science motivation had a positive effect on student achievement in science (*10 = .054, t(2.629); p < .001*), as did students' extrinsic science motivation, (*20 = .075, t(3.169); p < .01*). Importantly, students' dosage of the intervention (via membership in the classroom of a participating teacher) did not have a positive effect on student achievement. The sixth grade students had the least amount of "dosage" of the intervention (one year in the three-year intervention period).

For 7th graders, again using science benchmarks scores for outcome variable, both intrinsic science motivation (*10 = .051, t(2.694); p < .01*) and extrinsic motivation (*20 = .087 t(3.975); p < .001*) had positive effects on students' science achievement. An additional positive effect found in the 7th grade model for science achievement (with science benchmark serving as the outcome variable), was that of the dosage that students received of the intervention via being in a particular class with a participating teachers (*90 = .009, t(2.738); p < .01*). Science literacy also had a positive effect on student achievement (*130 = .126, t(4.868); p < .001*). Students' standardized English Language Arts (ELA) test scores were also positively related 7th grade students achievement in science (*140 = .070, t(36.638); p < .001*).

For 8th graders in science, again using science benchmarks scores as a measure of science achievement outcome variable, the results of the HLM conditional model indicate that intrinsic motivation ((*10 = .055, t(2.767); p < .05*) had a positive effect on student achievement, however extrinsic motivation had no effect on achievement, indicating that over time, and perhaps resulting from the intervention, the students became more intrinsically motivated, a goal of this intervention. Science reading comprehension (a sub factor of science literacy) had a positive effect on student achievement in 8th grade (*110 = .04,5 t(2.848); p < .05*). Student dosage of the intervention also had a positive effect on student achievement (*90 = .042, t(16.907); p < .001*). Finally, for 8th grade students, the HLM results for science indicate that students' ELA standardized test scores had a positive relationship with students' achievement in science (*140 = .06,5 t(36.51); p < .001*).

Additionally, for 8th grade students, using science standardized test (SST) scores as the outcome variable (a measure of science achievement; Note: this model could only be used for 8th grades as 6 and 7th graders do not take science SST), again students' intrinsic motivation had a positive effect on student science achievement (*10 = .358, t(2.215); p < .05*), similar to that of the results for the science benchmark outcome variable. Using the same 8th grade science HLM model, again science reading comprehension (a subset of science literacy) had a positive effect on students' achievement (*110 = .283, t(2.201); p < .05*). Once again, students ELA standardized test scores had a positive relationship with students' science achievement (*140 = .676, t(46.642); p < .001*). These results for 8th grade science achievement indicate that SST and science benchmarks scores are comparable as measures of 8th graders science achievement.

It is especially noteworthy that some of the precursing teacher factors (independent variables in Level 2 of the model) did not have a statistically significant effect on students' achievement in science. Teachers' years of experience and teachers' single subject credential (a proxy for pre intervention science or math content knowledge) did not have a statistically significant effect on students' achievement in science. This limited significance could be explained by the fact that content knowledge instruction and practice was embedded in each of the PD sessions over the three years, thereby increasing teachers' content knowledge while they were learning and practicing during the PD processes. Teachers' science teaching efficacy did not have a significant effect on students' achievement, except in the case of grade six where there was a modestly statistically significant effect.

In summary, students' motivation for science and their

		With Aggregate			Without Aggregate		
Model	Fixed Effects	Coeff.	t-ratio	p	Coeff.	t-ratio	p
Grade 6 Model	Model for mean SBK β_0						
	Intercept (γ_{00})	0.568	1.120	.292	0.431	2.700	.016
	TEACHER DOSE (γ_{01})	-0.065	-2.225	.053	0.004	0.167	.869
	TEACHER PERFORMANCE (γ_{02})	-0.015	-0.704	.500	0.012	0.504	.621
	TEACHER EFFICACY (γ_{03})	0.049	1.085	.306	0.092	2.225	.041
	TEACHER SINGLE CREDENTIAL (γ_{04})	0.020	0.561	.588	-0.005	-0.144	.888
	TEACHER EXPERIENCE (γ_{05})	0.004	0.299	.771	-0.017	-1.154	.265
	SCIENCE INTRINSIC MOTIVATION MEAN (γ_{06})	3.321	1.765	.111	-	-	-
	SCIENCE EXTRINSIC MOTIVATION MEAN (γ_{07})	2.554	1.149	.280	-	-	-
	SCIENCE MOTIVATION MEAN (γ_{08})	-6.162	-1.492	.170	-	-	-
	SCIENCE LITERACY WRITING MEAN (γ_{09})	-1.422	-0.890	.397	-	-	-
	SCIENCE LITERACY COMPREHENSION MEAN (γ_{10})	1.074	0.484	.640	-	-	-
	SCIENCE LITERACY VOCABULARY MEAN (γ_{11})	-1.330	-0.875	.404	-	-	-
	SCIENCE LITERACY TOTAL MEAN (γ_{12})	2.622	0.540	.602	-	-	-
	Model for Science Intrinsic Motivation Slope β_1						
	Intercept (γ_{10})	0.054	2.629	.009	0.054	2.633	.009
	Model for Science Extrinsic Motivation Slope β_2						
	Intercept (γ_{20})	0.075	3.169	.002	0.075	3.174	.002
	Model for Science Motivation Slope β_3						
	Intercept (γ_{30})	-0.102	-2.419	.016	-0.103	-2.423	.015
	Model for GENDER Slope β_4						
	Intercept (γ_{40})	0.008	2.050	.040	0.008	2.062	.039
	Model for HISPANIC Slope β_5						
	Intercept (γ_{50})	0.001	0.088	.930	0.002	0.109	.913
	Model for AFRICAN AMERICAN Slope β_6						
	Intercept (γ_{60})	-0.027	-1.695	.090	-0.027	-1.685	.092
	Intercept (γ_{70})	0.010	0.447	.655	0.010	0.453	.651
	Model for FILIPINO Slope β_8						
	Intercept (γ_{80})	0.036	1.363	.173	0.036	1.388	.165
	Model for STUDENT DOSE Slope β_9						
	Intercept (γ_{90})	-0.090	-10.559	.000	0.089	-10.436	.000
	Model for SCIENCE LITERACY WRITING Slope β_{10}						
	Intercept (γ_{100})	0.025	1.410	.159	0.025	1.403	.161
	Model for SCIENCE LITERACY COMPREHENSION Slope β_{11}						
	Intercept (γ_{110})	0.043	2.216	.027	.043	2.205	.028
	Model for SCIENCE LITERACY VOCABULARY Slope β_{12}						
	Intercept (γ_{120})	0.002	0.131	.896	0.002	0.124	.902
	Model for SCIENCE LITERACY TOTAL Slope β_{13}						
	Intercept (γ_{130})	0.091	1.887	.059	0.092	1.891	.059
	Model for ELA CST Slope β_{14}						
	Intercept (γ_{140})	0.062	29.119	.000	.062	29.151	.000
Grade 7 Model	Model for mean SBK β_0						
	Intercept (γ_{00})	0.383	0.513	.630	0.368	3.827	.002
	TEACHER DOSE (γ_{01})	0.004	0.090	.931	-0.011	-0.455	.657
	TEACHER PERFORMANCE (γ_{02})	-0.001	-0.047	.964	0.004	0.375	.714
	TEACHER EFFICACY (γ_{03})	0.046	0.708	.511	0.043	0.824	.426
	TEACHER SINGLE CREDENTIAL (γ_{04})	-0.047	-0.886	.416	-0.034	-1.073	.304
	TEACHER EXPERIENCE (γ_{05})	0.0004	0.029	.978	0.015	0.894	.389
	SCIENCE INTRINSIC MOTIVATION MEAN (γ_{06})	-.0113	-0.077	.942	-	-	-
	SCIENCE EXTRINSIC MOTIVATION MEAN (γ_{07})	0.297	0.170	.872	-	-	-
	SCIENCE MOTIVATION MEAN (γ_{08})	0.324	-0.105	.921	-	-	-
	SCIENCE LITERACY WRITING MEAN (γ_{09})	-0.335	-1.094	.324	-	-	-
	SCIENCE LITERACY COMPREHENSION MEAN (γ_{10})	0.936	2.236	.076	-	-	-
	SCIENCE LITERACY VOCABULARY MEAN (γ_{11})	0.729	0.903	.408	-	-	-
	SCIENCE LITERACY TOTAL MEAN (γ_{12})	-0.924	-0.631	.555	-	-	-
	Model for Science Intrinsic Motivation Slope β_1						
	Intercept (γ_{10})	0.051	2.694	.007	0.051	2.694	.007
	Model for Science Extrinsic Motivation Slope β_2						
	Intercept (γ_{20})	0.087	3.975	.000	0.087	3.978	.000
	Model for Science Motivation Slope β_3						
	Intercept (γ_{30})	-0.105	-2.680	.007	-0.105	-2.682	.007
	Model for GENDER Slope β_4						
	Intercept (γ_{40})	-0.002	-0.554	.579	-0.002	-0.554	.579
	Model for HISPANIC Slope β_5						
	Intercept (γ_{50})	-0.020	-1.401	.161	-0.020	-1.376	.169
	Model for AFRICAN AMERICAN Slope β_6						
	Intercept (γ_{60})	-0.045	-2.815	.005	-0.044	-2.802	.005
	Model for WHITE Slope β_7						
	Intercept (γ_{70})	-0.032	-1.560	.119	-0.032	-1.560	.119
	Model for FILIPINO Slope β_8						
	Intercept (γ_{80})	0.020	0.874	.382	0.020	0.882	.378
	Model for STUDENT DOSE Slope β_9						
	Intercept (γ_{90})	0.009	2.738	.006	0.008	2.597	.009
	Model for SCIENCE LITERACY WRITING Slope β_{10}						
	Intercept (γ_{100})	0.017	-1.995	.046	-0.017	-2.008	.045
	Model for SCIENCE LITERACY COMPREHENSION Slope β_{11}						
	Intercept (γ_{110})	-0.019	-1.317	.188	-0.020	-1.360	.174
	Model for SCIENCE LITERACY VOCABULARY Slope β_{12}						
	Intercept (γ_{120})	0.046	3.451	.001	0.045	3.431	.001
	Model for SCIENCE LITERACY TOTAL Slope β_{13}						
	Intercept (γ_{130})	0.126	4.868	.000	0.127	4.898	.000
	Model for ELA CST Slope β_{14}						
	Intercept (γ_{140})	0.070	36.638	.000	0.070	36.632	.000
Grade 8 Model	Model for mean SBK β_0						
	Intercept (γ_{00})	0.645	1.778	.136	0.507	7.279	.000
	TEACHER DOSE (γ_{01})	0.036	1.506	.192	0.026	1.307	.216
	TEACHER PERFORMANCE (γ_{02})	-0.013	-1.259	.264	-0.007	-0.747	.469
	TEACHER EFFICACY (γ_{03})	-0.001	-0.034	.974	0.024	0.731	.479
	TEACHER SINGLE CREDENTIAL (γ_{04})	-0.010	-0.208	.843	-0.022	-0.546	.595
	TEACHER EXPERIENCE (γ_{05})	-0.063	-2.374	.064	-0.033	-1.923	.078
	SCIENCE INTRINSIC MOTIVATION MEAN (γ_{06})	0.506	1.056	.339	-	-	-
	SCIENCE EXTRINSIC MOTIVATION MEAN (γ_{07})	0.644	-0.794	.463	-	-	-
	SCIENCE MOTIVATION MEAN (γ_{08})	0.053	0.064	.951	-	-	-
	SCIENCE LITERACY WRITING MEAN (γ_{09})	-0.396	-1.059	.338	-	-	-
	SCIENCE LITERACY COMPREHENSION MEAN (γ_{10})	0.913	1.314	.246	-	-	-
	SCIENCE LITERACY VOCABULARY MEAN (γ_{11})	0.488	0.575	.590	-	-	-
	SCIENCE LITERACY TOTAL MEAN (γ_{12})	-0.688	-0.470	.658	-	-	-
	Model for Science Intrinsic Motivation Slope β_1						
	Intercept (γ_{10})	0.055	2.767	.006	0.055	2.763	.006
	Model for Science Extrinsic Motivation Slope β_2						
	Intercept (γ_{20})	0.033	1.456	.145	0.033	1.448	.148
	Model for Science Motivation Slope β_3						
	Intercept (γ_{30})	-0.053	-1.306	.192	-0.053	-1.301	.193
	Model for GENDER Slope β_4						
	Intercept (γ_{40})	0.008	2.180	.029	0.008	2.165	.030
	Model for HISPANIC Slope β_5						
	Intercept (γ_{50})	-0.017	-1.228	.220	-0.017	-1.223	.221
	Model for AFRICAN AMERICAN Slope β_6						
	Intercept (γ_{60})	-0.040	-2.673	.008	-0.040	-2.674	.008
	Model for WHITE Slope β_7						
	Intercept (γ_{70})	-0.010	-0.513	.608	-0.010	-0.509	.611
	Model for FILIPINO Slope β_8						
	Intercept (γ_{80})	0.017	0.751	.453	0.017	0.743	.457
	Model for STUDENT DOSE Slope β_9						
	Intercept (γ_{90})	0.042	16.907	.000	0.042	16.978	.000
	Model for SCIENCE LITERACY WRITING Slope β_{10}						
	Intercept (γ_{100})	0.018	1.658	.097	0.018	1.656	.098
	Model for SCIENCE LITERACY COMPREHENSION Slope β_{11}						
	Intercept (γ_{110})	0.045	2.848	.004	0.045	2.842	.005

Table 5. Conditional Model using Science Benchmarks Scores (SBK) as Dependent Variable

Model	*Fixed Effects*	*Coefficient*	*t-ratio*	*p*	*Reliability*
Grade 8 Model	Intercept 1, β_0				.888
	Intercept 1, γ_{00}	3.745	41.118	.000	

Conditional Models (Science Standardized Test-SST-Scores as Dependent Variable)

Student – Level Model: (8)

$$\textbf{Science SST}_{ij} = \beta_{0j} + \beta_{1j} \times (\text{SCIENCE INTRINSIC MOTIVATION})_{ij} + \beta_{2j} \times (\text{SCIENCE EXTRINSIC MOTIVATION})_{ij} + \beta_{3j} \times (\text{SCIENCE MOTIVATION})_{ij} + \beta_{4j} \times (\text{GENDER})_{ij} + \beta_{5j} \times (\text{HISPANIC})_{ij} + \beta_{6j} \times (\text{AFRICAN AMERICAN})_{ij} + \beta_{7j} \times (\text{WHITE})_{ij} + \beta_{8j} \times (\text{FILIPINO})_{ij} + \beta_{9j} \times (\text{STUDENT DOSE})_{ij} + \beta_{10j} \times (\text{SCICE LITERACY WRITING})_{ij} + \beta_{11j} \times (\text{SCIENCE LITERACY COMPREHENSION})_{ij} + \beta_{12j} \times (\text{SCIENCE LITERACY VOCABULARY})_{ij} + \beta_{13j} \times (\text{SCIENCE LITERACY TOTAL})_{ij} + \beta_{14j} \times (\text{ELA SST})_{ij} + r_{ij}$$

Teacher – Level Model with the aggregate of variables: (9)

$$\beta_0 = \gamma_{00} + \gamma_{01} \times (\text{TEACHER DOSE})_j + \gamma_{02} \times (\text{TEACHER PERFORMANCE})_j + \gamma_{03} \times (\text{TEACHER EFFICACY})_j + \gamma_{04} \times (\text{TEACHER SINGLE CREDENTIAL})_j + \gamma_{05} \times (\text{TEACHER EXPRIENCE})_j + \gamma_{06} \times (\overline{\text{SCIENCE INTRINSIC MOTIVATION}})_j + \gamma_{07} \times (\overline{\text{SCIENCE EXTRINSIC MOTIVATION}})_j + \gamma_{08} \times (\overline{\text{SCIENCE MOTIVATION}})_j + \gamma_{09} \times (\overline{\text{SCIENCE LITERACY WRITING}})_j + \gamma_{10} \times (\overline{\text{SCIENCE LITERACY COMPREHENSION}})_j + \gamma_{11} \times (\overline{\text{SCIENCE LITERACY VOCABULARY}})_j + \gamma_{12} \times (\overline{\text{SCIENCE LITERACY TOTAL}})_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}$$
$$\beta_{6j} = \gamma_{60}$$
$$\beta_{7j} = \gamma_{70}$$
$$\beta_{8j} = \gamma_{80}$$
$$\beta_{9j} = \gamma_{90}$$
$$\beta_{10j} = \gamma_{100}$$
$$\beta_{11j} = \gamma_{110}$$
$$\beta_{12j} = \gamma_{120}$$
$$\beta_{13j} = \gamma_{130}$$
$$\beta_{14j} = \gamma_{140}$$

Teacher – Level Model without the aggregate of variables: (10)

$$\beta_0 = \gamma_{00} + \gamma_{01} \times (\text{TEACHER DOSE})_j + \gamma_{02} \times (\text{TEACHER PERFORMANCE})_j + \gamma_{03} \times (\text{TEACHER EFFICACY})_j + \gamma_{04} \times (\text{TEACHER SINGLE CREDENTIAL})_j + \gamma_{05} \times (\text{TEACHER EXPRIENCE})_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}$$
$$\beta_{9j} = \gamma_{90}$$

SCIENCE INTRINSIC MOTIVATION, SCIENCE EXTRINSIC MOTIVATION and SCIENCE MOTIVATION have been centered around the group mean.

TEACHER EFFICACY has been centered around the grand mean.

Table 6. Unconditional Model using Science Standardized Test (SST) Scores as DV

science literacy had a positive effect on their achievement. Further, the dosage of students' exposure to the teachers' PD had a positive effect on student achievement.

Discussion

As previously stated, the purpose of this research was to investigate the impact of a science teacher professional development intervention on middle school teachers' efficacy in teaching science, and ultimately, on their students' achievement, literacy in and motivation for science. The intervention was found to be highly effective in positively impacting students' achievement, motivation and interest in science. The results of this research indicate that both teacher pre-intervention factors and school level factors did not significantly impact student achievement in the subject area under study. This supports the understanding that the intervention itself impacted student achievement across the three intervention years. The HLM results confirmed and explained the reasons for school-wide achievement and academic performance indices (API) gains for the five participating schools. Specifically, the results indicate that increases in student dosage of the intervention (by proxy of their teachers' participation in the intervention) led to student gains in achievement, thus confirming that the intervention positively impacted student achievement; the primary goal of the intervention. Given that the intervention was designed to impact both students' achievement in science and their interest in and motivation for science, by deliberately instructing and guiding teachers on how to use science experimentation in their class as a means of increasing students' interest, motivation for and achievement in the subject, and a target was placed on literacy in science as a means of positively impacting science achievement by providing students with tools for effectively using and comprehending their science texts, the intervention was highly successful for the participating students. The participating teachers also increased their science teaching efficacy over time, another project goal. While the teaching efficacy did not greatly impact the students' achievement in our particular study, given its relevance in teacher professional development literature, we believe it was an important outcome for the teachers who participated with us.

The results of this research are highly generalizable to others who might attempt this structure and type of intervention. First, it was built upon existing, impactful research as described in the literature review section of this paper. Second, both the teacher and student populations under study mirror that of many urban middle schools nationally (also described in the review of the literature). Third, given that the study results indicate that school level effects (via the three-level HLM model), did not predict the achievement of the students in the study, the results suggest that the comprehensiveness of the intervention could apply to diverse school settings.

There are several factors that contributed to the intervention's effects and impacts. The project team achieved maximum fidelity of implementation at the teacher level of the intervention because of the structure employed to guide and implement the teacher professional development. Additionally, using data driven decision-making and using student achievement data as a metric for intervention content maximized the potential content impact for the students in the intervention schools. Finally, the fact that there was no teacher attrition in the research

Model	Fixed Effects	With Aggregate Coeff.	With Aggregate t-ratio	With Aggregate p	Without Aggregate Coeff.	Without Aggregate t-ratio	Without Aggregate p
Grade 8 Model	Model for mean Science SST β_0						
	Intercept (γ_{00})	5.302	2.625	.047	1.680	5.170	.000
	TEACHER DOSE (γ_{01})	0.329	2.500	.054	0.149	1.725	.110
	TEACHER PERFORMANCE (γ_{02})	-0.003	-0.051	.961	0.008	0.182	0.859
	TEACHER EFFICACY (γ_{03})	0.338	1.428	.213	0.061	0.455	.657
	TEACHER SINGLE CREDENTIAL (γ_{04})	-0.181	-0.685	.524	0.030	0.173	.866
	TEACHER EXPERIENCE (γ_{05})	-0.437	-2.984	.031	-0.255	-3.450	.005
	SCIENCE INTRINSIC MOTIVATION MEAN (γ_{06})	3.330	1.224	.275	-	-	-
	SCIENCE EXTRINSIC MOTIVATION MEAN (γ_{07})	4.130	0.901	.409	-	-	-
	SCIENCE MOTIVATION MEAN (γ_{08})	-8.042	-1.648	.160	-	-	-
	SCIENCE LITERACY WRITING MEAN (γ_{09})	-1.828	-0.883	.418	-	-	-
	SCIENCE LITERACY COMPREHENSION MEAN (γ_{10})	0.573	0.147	.889	-	-	-
	SCIENCE LITERACY VOCABULARY MEAN (γ_{11})	5.506	-1.152	.301	-	-	-
	SCIENCE LITERACY TOTAL MEAN (γ_{12})	5.098	0.622	.561	-	-	-
	Model for Science Intrinsic Motivation Slope β_1						
	Intercept (γ_{10})	0.358	2.215	.027	0.357	2.211	.027
	Model for Science Extrinsic Motivation Slope β_2						
	Intercept (γ_{20})	0.307	1.674	.094	0.307	1.674	.094
	Model for Science Motivation Slope β_3						
	Intercept (γ_{30})	-0.536	-1.620	.105	-0.535	-1.618	0.106
	Model for GENDER Slope β_4						
	Intercept (γ_{40})	0.286	10.060	.000	0.286	10.075	.000
	Model for HISPANIC Slope β_5						
	Intercept (γ_{50})	-0.079	-0.701	.483	-0.080	-0.713	.476
	Model for AFRICAN AMERICAN Slope β_6						
	Intercept (γ_{60})	-0.259	-2.118	.034	-0.262	-2.140	.032
	Model for WHITE Slope β_7						
	Intercept (γ_{70})	0.094	0.603	.546	0.093	0.599	.549
	Model for FILIPINO Slope β_8						
	Intercept (γ_{80})	0.139	0.757	.449	0.138	0.752	.452
	Model for STUDENT DOSE Slope β_9						
	Intercept (γ_{90})	0.027	1.343	.179	0.026	1.290	.197
	Model for SCIENCE LITERACY WRITING Slope β_{10}						
	Intercept (γ_{100})	0.013	0.145	.885	0.012	0.134	.894
	Model for SCIENCE LITERACY COMPREHENSION Slope β_{11}						
	Intercept (γ_{110})	0.283	2.201	.028	0.282	2.194	.028
	Model for SCIENCE LITERACY VOCABULARY Slope β_{12}						
	Intercept (γ_{120})	-0.010	-0.090	.928	-0.010	-0.091	.928
	Model for SCIENCE LITERACY TOTAL Slope β_{13}						
	Intercept (γ_{130})	0.355	1.431	.152	0.356	1.434	.152
	Model for ELA CST Slope β_{14}						
	Intercept (γ_{140})	0.676	46.642	.000	0.676	46.748	.000

MODEL	Random Effects	With Aggregate Variance	With Aggregate Df	With Aggregate Chi-square (p)	Without Aggregate Variance	Without Aggregate df	Without Aggregate Chi-square (p)
Grade 6 Model	Intercept 1, u_0 Level -1, r	0.003 0.011	9	333.384 ($p < .001$)	0.005 0.011	16	935.160 ($p < .001$)
Grade 7 Model	Intercept 1, u_0 Level -1, r	0.002 0.009	5	124.309 ($p < .001$)	0.003 0.009	12	389.332 ($p < .001$)
Grade 8 Model	Intercept 1, u_0 Level -1, r	0.002 0.009	5	143.685 ($p < .001$)	0.002 0.009	12	544.060 ($p < .001$)

Table 7. Conditional Model using Science Standardized Test (SST) Scores as Dependent Variable

led to maximum doses of intervention by year three of the intervention with the 8th grade students. While attrition was minimal for this study, it is a national issue. We recognize this difficulty and believe it is helpful to have cohort programs in teacher PD, as this has been found to lead to less attrition as teachers are able to support one another in their cohorts, especially if they engage in lesson study as a part of their ongoing professional development.

Study Limitations

The study is limited by several factors. Teachers were the primary administers of the student outcome measurements and therefore were responsible for collecting all student level data. This was only modestly problematic because the research team optimized the structure for administration and data collection, made it similar to that of standardized K-12 assessment procedures, and kept these structure constant semester-to-semester throughout the study. Additionally, the design of the intervention did not allow for pre-post comparisons of teacher performance because lesson study was the means by which the TPOR scores were obtained and the participating teachers only submitted one teaching video (used to assess performance using the TPOR) during the intervention time frame to study. This related to two factors: (a) the participating teachers needed to learn how to engage in lesson study as a critical structure for the intervention during which the videos were produced; and (b) the three-year intervention period did not allow for substantial time to videotape each teacher twice for instructional performance comparison purposes.

Teacher attrition was not a challenge in the intervention in spite of the socio-political climate of K-12 public schools associated with the nation's economic difficulties during the intervention period. The participating school district made a commitment to retain all science and math teachers during the intervention period (1) because the need for stability of teachers in the subject area was great, considering the academic needs of the students, and (2) the district recognized the importance of stability of teacher participants for the research funding period. The district's teachers' union and administration supported this decision at all levels.

Implication for Research

The findings of this study are highly significant to researchers. The results provide ample evidence that teacher professional development can positively impact student achievement while simultaneously positively impacting teacher instructional performance and teaching efficacy. Researchers in higher education have been highly skeptical about the impact of in-service teacher interventions on students' achievement. This research negates such skepticism, especially given its scale.

Research points to three elements that are crucially important for effective teaching: (1) teachers' deep knowledge of the subject matter to be taught (or teacher content knowledge), (2) skill in how to teach the subject, and (3) an understanding of how people learn (Darling-Hammond, 2000; Donovan & Bransford, 2005; Weston, Hindley, & Cunningham, 2021). While the latter two elements are necessary and important, they alone cannot determine whether the teacher is able to teach so that their students acquire a deep understanding of science content. This study provides empirical evidence to support educational researchers to continue emphasizing the importance of content knowledge in professional development for teachers to bring about change in science teaching practice and in student learning and achievement.

Implication for Field

Being supported by existing, impactful research as described in the literature review section of the paper and the findings from the HLM model, the intervention can be applied to diverse school settings to address the critical in-service teacher PD needs that have been underscored by research by Nye, Konstantopoulos, and Hedges (2004), Darling-Hammond, (2012), and associated PD research by Berry (2010), Clotfelter and colleagues, (2006) and Davidson and Hughes (2018). The following suggestions are made to optimize the practical application of the intervention designed in this study.

First, involve teachers and middle school administration in the iterative development process of the intervention and identify contemporary teacher and student needs to facilitate "buy-in" from the teachers and school administrators. Second, because the intensive portions of the intervention were designed to occur independently of classes, the teachers and administrators can start taking/implementing the intervention in summer so that their teaching contract and schedule during the academic years will be minimally interrupted. Third, the teachers can continuously improve their content knowledge by forming disciplinary communities with peers and university faculty (Weston, Hindley, & Cunningham, 2021). The teachers are encouraged to communicate and collaborate with university experts via content mentorship and face-to-face follow-up lesson study as described in this study. Finally, these interventions should also be utilized in preservice programs as a means of proactively influencing teachers and ultimately their students.

Authors' Note

This research was supported by funding provided by a state department of education. The research team for this project would like to thank the leadership of this state department for their support and guidance throughout this important teacher professional development and child intervention project. The team would also like to thank the school district leadership and the five school site principals for their guidance and engaged support for this research endeavor. A special thanks also belongs to the middle school teachers for their full participation in the professional development and research associated with the project and for their heartfelt perseverance with their 6th-8th grade students during the project's three intervention years. A final thanks belongs to the school district's middle school students and their families for participating in this important school-based intervention effort.

References

Akerson, V.L., Pongsanon, K., Park Rogers, M.A., Carter, I. & Galindo, E. (2015). Exploring the use of lesson study to develop elementary preservice teachers' pedagogical content knowledge for teaching nature of science, *IJSME, 15*(2), 293–312. https://doi.org/10.1007/s10763-015-9690-x

Ambussaidi, I., & Yang, Y. F. (2019). The impact of mathematics teacher quality on student achievement in Oman and Taiwan. *IJELE, 1*(2), 50-62. https://doi.org/10.31763/ijele.v1i2.39

Anthony, R.J., Tippett, C.D. & Yore, L.D. (2010). Pacific CRYSTAL project: Explicit literacy instruction embedded in middle school science classrooms. *RISE, 40*(1), 45–64. https://doi.org/10.1007/s11165-009-9156-7

Ball, D. L., & Forzani, F. M. (2010, December 1). *Teaching skillful teaching. ASCD*. https://www.ascd.org/el/articles/teaching-skillful-teaching

Burns, M.K., Naughton, M.R., Preast, J.L., Wang, Z., Gordon, R.L., Robb V. & Smith, M.L. (2018) Factors of professional learning community implementation and effect on student achievement, *JEPC, 28*(4), 394-412. http://doi.org/10.1080/10474412.2017.1385396

Bybee, R.W. (2010). Advancing STEM education: A 2020 vision. *TET, 70*(1), 30–35.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Chan, K. H. K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran J., Mavhunga, E., Neumann, K., ... Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp.77-94). Springer, Singapore. https://doi.org/10.1007/978-981-13-5898-2_2

Cajkler, W. & Wood, P. (2016) Adapting 'lesson study' to investigate classroom pedagogy in initial teacher education: what student-teachers think, *Cambridge Journal of Education, 46*(1), 1-18, http://doi.org/10.1080/0305764X.2015.1009363

California Department of Education (2021, June 12). *Taking Center Stage-Act II: Ensuring Success and Closing the Achievement Gap for All of the State's Middle Grades Students.* Report to the Department of Education. https://www.cde.ca.gov/ci/gs/mg/tcsii-index.asp

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *JHR, XVL*(4), 778–820. https://doi.org/10.3368/jhr.XLI.4.778

Cohen, J. (1992). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.https://doi-org.libproxy1.usc.edu/10.4324/9780203771587

Darling-Hammond, L. (2012). Why teacher education is important and – Difficult. In *Powerful teacher education: Lessons from exemplary programs* (pp. 121-154). John Wiley & Sons.

Darling-Hammond, L. (2000). How teacher education matter. *JTE, 51*(3), 161–173. https://doi.org/10.1177/0022487100051003002

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *EPAA*, 8(1), 1–44. https://doi.org/10.14507/epaa.v8n1.2000

Davidson, S.G., & Hughes, R. (2018). Communities of practice to explain teachers' experiences within the community of science. *JRST, 55*(9), 1287-1312. https://doi.org/10.1002/tea.21452

Donovan, M. S., & Bransford, J.D. (Eds.). (2005). How students learn: History, mathematics, and science. National Academies Press. https://doi.org/10.17226/10126

Fang, Z., & Wei, Y. (2010). Improving middle school students' science literacy through reading infusion. The *Journal of Educational Research, 103*(4), 262–273. https://doi.org/10.1080/00220670903383051

Gerard, L. F., Varma, K., Corliss, S. B., & Linn, M. C. (2011). Professional development for technology-enhanced inquiry science. *RER, 81*(3), 408– 448. https://doi.org/10.3102/0034654311415121

Granger, E.M., Bevis, T.H., Southerland, S.A., Saka, Y., & Ke, F. (2018). Examining features of how professional development and enactment of educative curricula influences elementary science teacher learning. *JRST, 56*(3), 348 – 370. https://doi.org/10.1002/tea.21480

Guthrie, J. T., McRae, A., & Klauda, S. L. (2007). Contributions of concept-oriented reading instruction to knowledge about interventions for motivations in reading. *Educational Psychologist, 42*(4), 237 – 250. https://doi.org/10.1080/00461520701621087

Jones, M. H., & Gallen, A. M. (2016). Peer observation, feedback and reflection for development of practice in synchronous online teaching. *Innovation in Education & Teaching International, 53*(6), 616-626. https://doi.org/10.1080/14703297.2015.1025808

Kanelloupou, E. M., & Darra, M. (2018a). The contribution of lesson study to the development of a collaborative framework for the operation of the school: Results from a pilot implementation in secondary education in Greece. *IJCER, 1*(2), 33-45. https://doi.org/10.11114/ijce.v1i2.3603

Kanellopoulou, E. M., & Darra, M. (2018b). The implementation of the lesson study approach to secondary education in Greece: The case of the literature lesson. *IJLTER, 17*(7), 94-105. https://doi.org/10.26803/ijlter.17.7.6

Kayapinar, U. (2016). A study on reflection in in-service teacher development: Introducing reflective practitioner development model. *Educational Sciences: Theory & Practice, 16*(5), 1671-1691. https://doi.org/10.12738/estp.2016.5.0077

Keller, M.M., Hoy, A.W., Goetz, T., & Frenzel, A.C. (2016). Teacher enthusiasm: Reviewing and redefining a complex construct. *Educational Psychology Review, 28*(4), 743–769. https://doi.org/10.1007/s10648-015-9354-y

Kohlmeier, J., Howell, J., Saye, J. McCormick, T., Shannon, D., Jones, C. & Brush, T. (2020) Investigating teacher adoption of authentic pedagogy through lesson study. *TRSE, 48*(4), 492–528. https://doi.org/10.1080/00933104.2020.1751761

Kolleck, N. (2019). Motivational aspects of teacher collaboration. *Frontiers in Education, 4,* https://doi.org/10.3389/feduc.2019.00122

Lazarides, R., Fauth, B., Gaspard, H., & Göllner, R. (2021). Teacher self-efficacy and enthusiasm: Relations to changes in student-perceived teaching quality at the beginning of secondary education. *Learning & Instruction, 73*, 101435. https://doi.org/10.1016/j.learninstruc.2020.101435

Lewis, C.C., Perry, R.R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *JMTE, 12*(4), 285–304. https://doi.org/10.1007/s10857-009-9102-7

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry.* Sage.

Listyani, E., Widjajanti, D. B., Susanti, M., Arliani, E., & Hidayati, K. (2008). Development of mathematics high school teachers' competency through lesson study (A case study in Yogyakarta, Indonesia). International Conference on Lesson Study. 1–9.

Mahan, K. R. (2020). The comprehending teacher: Scaffolding in content and language integrated learning (CLIL). LLJ, 48(1), 1-15. https://doi.org/10.180/09571736.2019.1705879

McConnell, T.J., Parker, J.M., & Eberhardt, J. (2013). Assessing teachers' science content knowledge: A strategy for assessing depth of understanding. *JSTE, 24*(4), 717–743. https://doi.org/10.1007/s10972-013-9342-3

McNally, J.C. (2015). Learning from one's own teaching: New science teachers analyzing their practice through classroom observation cycles. *JRST, 53*(3), 473-501. https://doi.org/10.1002/tea.21253

Mundry, S. (2005). Changing perspectives in professional development. *Science Educator, 14*(1), 9–15.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter future.* National Academies Press. https://doi.org/10.17226/11463

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *EEPA, 26*(3), 237–257. https://doi.org/10.3102/01623737026003237

Ostovar-Nameghi, S. A., & Sheikhahmadi, M. (2016). From teacher isolation to teacher collaboration: Theoretical perspectives and empirical findings. *ELT, 9*(5), 197-205. http://dx.doi.org/10.5539/elt.v9n5p197

Parker, J.M, McConnell, T.J., & Eberhardt, J. (2013). *Characterizing teachers' incoming science content knowledge in a professional development program* [Paper]. NARST 2013 Conference, Rio Grande, Puerto Rico.

Ragusa, G. (2011). *Best Practices in Pedagogy and Associated Assessment* [Presentation]. National Science Foundation EEC Awardees, Arlington, VA.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Sage.

Riggs, I.M. & Enochs, L.G. (1990) Toward Development of an Elementary Teachers' Science Teachers Efficacy Belief Instrument. *Science Education 74*(6), 625-637.

Rozimella, Y. (2020). Developing teachers' professionalism through school-initiative lesson study. *EU-JER, 9*(4), 1513-1526. https://doi.org/10.12973/EU-JER.9.4.1513

Sadler. P.M., Sonnert, G., Coyle, H.P., Cook-Smith, N., & Miller, J.L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *AERJ, 50*(2), 1–30. https://doi.org/10.3102/0002831213477680

Saito, E., & Atencio, M. (2013). A conceptual discussion of lesson study from a micropolitical perspective: Implications for teacher development and pupil learning. *Teaching & Teacher Education, 31*(1), 87–95. https://doi.org/10.1016/j.tate.2013.01.001

Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of lesson study. Teaching & Teacher Education, 68, 289–303. https://doi.org/10.1016/j.tate.2017.09.015

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *ER, 15*(2), 4-31. https://doi.org/10.3102/0013189X015002004

Stigler, J. W., & Hiebert, J. (2016). Lesson study, improvement, and the importing of cultural routines. *Mathematics Education, 48*(4), 581-587. https://doi.org/10.1007/s11858-016-0787-7

Stigler, J. (2006). Lesson study: Principles and practices. *Teacher Education, 97*(3), 43–61.

Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (2015). An efficacy trial of research-based curriculum materials with curriculum-based professional development. *AERJ, 52*(5), 984–1017. https://doi.org/10.3102/0002831215585962

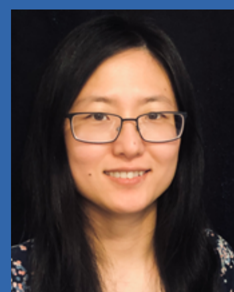
Tretter. T.R., Brown, S.L., Bush, W.S., Saderholm, J., & Holmes, V.-L. (2013). Valid and reliable science content assessments for science teachers. *JSTE, 24*(2), 269–295. https://doi.org/10.1007/s10972-012-9299-7

Van Driel, J., Jonh, O. D., & Verloop, N. (2021). The development of preservice chemistry teachers' pedagogical content knowledge. *Science Education, 86*(4), 572-590. https://doi.org/10.1002/sce.10010

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review, 15*, 17-40. https://doi.org/10.1016/j.edurev.2015.04.002

Dr. Gisele Ragusa received her Ph.D. from the University of Southern California in Los Angeles, California. She is currently a Professor of Engineering Education at the University of Southern California. Before becoming a professor, she was a K-12 teacher for 14+ years. Her research interests include science, technology, engineering and mathematics (STEM) education, engineering innovation, college access, STEM PK-12 education and teacher education, student mentorship, STEM literacy education, human-centered design, socially assistive robotics, as well as assessment and measurement. Dr. Ragusa has received the US Presidential Award for Excellence in Science, Mathematics, and Engineering Mentoring (PAESMEM) Award in 2019.

Dr. Shaobo Huang received a Ph.D. degree in Engineering Education from Utah State University and collaborated with other authors on this paper as a postdoctoral associate researcher at the University of Southern California. She has over eight years of teaching and/or research experience in engineering education. She is currently an Assistant Professor in the Ron and Jane Graham School of Professional Development with a joint appointment in the Department of Mechanical Engineering at the University of Saskatchewan in Canada. Dr. Huang is one of the key leaders on the RE-ENGINEERED First-Year Program at the University of Saskatchewan.



Dr. Svetlana V. Levonisova earned her Ed.D. degree in Administration in Higher Education from the University of Southern California. She participated in this research paper as a postdoctoral scholar. She is an international educator with 20 years of experience in higher education including teaching and research that she started in Russia. Currently, she is an Assessment Coordinator at the Attallah College of Educational Studies at Chapman University. She is working for the Office of Accreditation and Assessment to coordinate and support state and national accreditation processes for seven graduate and one undergraduate programs.