Research Article

The Impact of a *Framework*-Aligned Science Professional Development Program on Literacy and Mathematics Achievement of K-3 Students

Peter Paprzycki,[1] Nicole Tuttle ([iD]),[2] Charlene M. Czerniak,[2] Scott Molitor,[2] Joan Kadervaek,[3] and Robert Mendenhall[4]

[1]*Department of Educational Research and Administration, University of Southern Mississippi, Hattiesburg, Mississippi*
[2]*College of Engineering, University of Toledo, Toledo, Ohio*
[3]*Department of Early Childhood, Physical and Special Education, University of Toledo, Toledo, Ohio*
[4]*Toledo Public Schools, Toledo, Ohio*

Abstract: This study investigates the effect of a *Framework*-aligned professional development program at the PreK-3 level. The NSF funded program integrated science with literacy and mathematics learning and provided teacher professional development, along with materials and programming for parents to encourage science investigations and discourse around science in the home. This quasi-experimental study used a three-level hierarchical linear model to compare the Renaissance STAR Early Literacy, Reading, and Mathematics scores from 2015 to 2016 of K-3 students in treatment and control classrooms in a large Midwestern urban school district. The statistically significant results indicate that, on average, every year that a student has a program teacher adds 8.6 points to a student's spring STAR Early Literacy score, 17.0 points to a student's STAR Mathematics score, and 41.4 points to a student's STAR Reading score compared to control students. Implications for early elementary teacher education and policy are discussed. © 2017 Wiley Periodicals, Inc. J Res Sci Teach 9999:XX–XX, 2017

Keywords: language and literacy; mathematics; professional development

Science reform efforts face many challenges. At the elementary level, a major challenge is the lack of time devoted to teaching science (Blank, 2013). The pressure of high-stakes testing can prompt schools to reduce instructional time in science, while increasing time spent on reading and mathematics (McMurrer, 2007; Milner et al., 2012). Younger students in particular receive little science instruction. Numerous studies have demonstrated that early childhood educators provide little science instruction (Nayfeld, Brenneman, & Gelman, 2011; Piasta, Yeager Pelatti, & Miller, 2014; Tu, 2006). Time-use studies have shown that early childhood educators often stress language and literacy learning more heavily than math or science (Early et al., 2010;

La Paro et al., 2009). Furthermore, time spent in increased reading instruction often focuses on basal reading with little emphasis on content learning (Jones et al., 1999).

Such a paradigm conflicts with years of research emphasizing the importance of integrating science learning with reading and mathematics. A mixed-methods meta-analysis of studies that examined the impact of integrated science and mathematics programs from Hurley (2001) looked at 31 studies and found that most supported integration. Reading researchers have long advocated linking literacy with science instruction (Chall & Jacobs, 2003; Guthrie & Ozgungor, 2002; Pearson, Moje, & Greenleaf, 2010; Snow, 2002). And in the United States, the most recent reform efforts, *A Framework for K-12 Science Education* (*Framework*) (National Research Council [NRC], 2012) and the *Next Generation Science Standards* (NGSS) (NGSS Lead States, 2013), provide a foundation for effective science learning that emphasizes integrating science inquiry with reading and mathematics skills. In this model, science instruction aligned to the *Framework* asks students to investigate natural phenomena, read, discuss, and write about their investigations, use mathematics and computational thinking to analyze their data and draw conclusions, and make arguments from evidence, just as professional scientists carry out their work. The *Framework* particularly focuses on the importance of language for students' science knowledge development: "Any education in science and engineering needs to develop students' ability to read and produce domain-specific text. As such, every science or engineering lesson is in part a language lesson, particularly reading and producing the genres of texts that are intrinsic to science and engineering" (NRC, 2012, p. 76).

The *Framework*, therefore, represents a new model for science instruction integrating reading and mathematics. In this study, we examine the effect of a *Framework*-aligned science professional development [PD] program, NURTURES, on K-4 student achievement in early literacy, reading, and mathematics. The NURTURES PD placed special emphasis on elevating teachers' emphasis on incorporating expository text into science instruction, fostering classroom discourse (including science vocabulary and inferential thinking), along with a focus on students' mathematical thinking. This study, therefore, contributes to research demonstrating the effects of science/reading and science/math-integrated curricula by extending it to a novel model for integration, the *Framework* (NRC, 2012), and to the early elementary, which is an under-studied area of science education.

Need for Science PD in Early Elementary Grades

Early childhood educators report feeling particularly underprepared to teach science (Greenfield et al., 2009). One reason for this is their limited exposure to science instructional practices during teacher preparation programs (Lobman, McLaughlin, & Ryan, 2005) and during PD (Wilson, Lubienski, & Mattson, 1996; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In their 2012 review research on science inquiry PD, Capps, Crawford, and Constas found published research from just 17 of these PD programs, with almost all of those works focused on late elementary or secondary teachers. Furthermore, a recent study demonstrated that the science achievement gap begins in kindergarten and persists at least to eighth grade (Morgan, Farkas, Hillemeier, & Maczuga, 2016), speaking to the need for high quality science instruction to begin in the earliest grades.

Recent results from the NURTURES program examined whether *Framework*-aligned PD could influence early elementary teachers to aligned their science instruction to the *Framework* (Tuttle et al., 2016). Changes in science content knowledge, lesson plans, and classroom practices were examined from before and after the 2014 NURTURES PD in 11 teachers from grades 2 to 3. The study found that teachers increased their overall use of science inquiry practices, particularly several related to classroom discourse. It also noted that the teachers' lesson plans indicated that

they were incorporating science expository texts, science vocabulary, and mathematical thinking into their science instruction.

Links Between Science, Reading, and Mathematics

Multiple lines of reasoning contribute to the research basis for integrating science and reading instruction. Content knowledge, including science content knowledge, is critical to strengthening reading skills. For example, there is a strong link between knowledge of vocabulary and reading achievement (National Reading Panel, 2000). Furthermore, it has been demonstrated that argumentation, critical reading, and writing are critical for promoting science literacy (e.g., Glynn & Muth, 1994; Holliday et al., 1994; Shymansky et al., 2000; Yore et al., 2004). Finally, research that evaluates the impact of merging science and literacy instruction has established that such approaches benefit both science and literacy learning (e.g., Cervetti, Barber, Dorph, Pearson, & Goldschmidt, 2012; Hapgood & Palincsar, 2007; Palincsar & Magnusson, 2001; Romance & Vitale, 1992, 2001; Varelas & Pappas, 2006).

Although the research establishing clear benefits for integrating reading and math are well established upper elementary and middle school students, there are far fewer studies examining integrated instruction in the early elementary years. One example of science-literacy connections in early elementary is the work of Varelas and coworkers, who have studied the opportunities to develop Latinx students' science understanding afforded by read-alouds of science information books and related hands-on explorations (Varelas & Pappas, 2006; Varelas, Pappas, & Rife, 2006; Varelas, Pieper, Arsenault, Pappas, & Keblawe-Shamah, 2014). For example, in a study of first and second graders, Varelas and Pappas (2006), teachers read aloud from and discussed books about the water cycle and states of matter, providing opportunities for students to make connections to their personal experiences and the texts. Students also participated in related hands-on science investigations. Over time, students developed more connections among their experience, the texts, and the investigations, used the vocabulary from the texts, and began to develop their own ideas about how the world works.

In another study, the IDEAS program (Romance & Vitale, 1992, 2001, 2011, 2012) replaces language arts instruction with a joint science-reading program in participating schools. The IDEAS model incorporates direct instruction, science investigations, reading, writing, and concept mapping during this joint instruction time. A number of studies demonstrated that students in IDEAS classrooms outperform control students on standardized assessments of science and reading achievement. In a study of students in grades 1–2, they found that when implementing an IDEAS model that incorporated the instruction into daily 45-minute blocks rather than completely replacing reading instruction, students in IDEAS classrooms again outperformed control students on the Iowa Test of Basic Skills science and reading assessments.

The *Framework* (NRC, 2012) also emphasizes the role of mathematics in science instruction: "Increasing students' familiarity with the role of mathematics in science is central to developing a deeper understanding of how science works (NRC, 2012, p. 66). Using mathematics and computational thinking is one of the eight science and engineering practices of the *Framework* (NRC, 2012), and a second, analyzing and interpreting data, echoes the "Measurement and Data" domain of the Common Core state mathematics standards (National Governors Association Center for Best Practices, 2010). The quantitative analysis indicated that integration enhanced both mathematics and science achievement, though the method of integration determined the extent of the effects. When mathematics was used to enhance science or when the two subjects were fully integrated, science achievement gains tended to be higher than mathematics achievement gains. Conversely, in studies where mathematics was planned in sequence with science, but the subjects were taught separately, mathematics achievement gains were higher.

More recently, a study of a STEM enrichment program demonstrated that middle school students who participated realized statistically significant growth on state assessments of science and mathematics, compared to comparison students from the same district (Sondergeld, Milner, Coleman, & Southern, 2011).

Cross Talk Among Science, Reading, and Mathematics

The *Framework*-aligned program examined in this study, due to its emphasis on using the science and engineering practices (NRC, 2012), promotes science instruction that emphasizes inquiry investigations and also integrates literacy and mathematics instruction. This novel model for science and literacy integration draws on common practices for instruction in the three disciplines that can work together synergistically to enhance student learning in all arenas. We argue that the parallel skills of reading, mathematics, and science may allow gains in science learning to accompany gains in reading and mathematics achievement.

Importantly, both science and reading learning benefit from student discourse. Spoken language is an important tool to help students use scientific reasoning (Mercer, Dawes, Wegerif, & Sams, 2004) and make their scientific thinking "visible" (Newton, Newton, Blake & Brown, 2002; Varelas & Pappas, 2006), along with supporting language and literacy development (Romance & Vitale, 2001, 2011; Krajcik & Sutherland, 2010). Specifically, student discourse during science can support inferential thinking skills (Zucker, Justice, Piasta, & Kaderavek, 2010), which are an important precursor to reading comprehension (van Kleeck, Vander Woude, & Hammett, 2006).

A second area for cross-pollination between reading and science instruction is teaching and using science vocabulary. As previously noted, knowledge of vocabulary and reading achievement are strongly correlated (National Reading Panel, 2000). As Varelas et al. (2014) points out, students' vocabularies expand when they are introduced to terms, encounter scientific terms in context, and explore and discuss science vocabulary during science investigations. Bravo and Cervetti (2008) further argue that learning science vocabulary is science content learning. When students make connections among science terms, it leads to the formation of rich conceptual networks. Building science vocabulary, therefore, both contributes to students' science learning but also enhances their ability to read science texts.

Teaching with science informational texts is another area of synergy between reading and science in an integrated curriculum. Science instruction offers a critical opportunity to foster children's early literacy experiences because when children are engaged with others in a science inquiry investigation, they are active participants engaged in a community of practice using literacy tools to jointly investigate, discuss, research, and document the outcomes of their investigations (Wenger & Lave, 1991). This is important because science inquiry provides a context for literacy learning, rather than a decontextualized experience where children are expected to exhibit isolated literacy skills, and because a science inquiry investigation can motivate children to use literacy when it is a part of the exploratory process. Increased levels of motivation to read and write have been demonstrated to improve literacy outcomes (Wigfield & Guthrie, 2000). Furthermore, research in educational psychology provides the theory that active participation and manipulation of objects described in texts improves children's ability to mentally map words to text thus improving reading comprehension (e.g., Glenberg, 2011).

Calls to integrate mathematics and science date back many decades (Breslich, 1936). A number of evidence-based benefits to integrating mathematics and science exist (Czerniak & Johnson, 2014). Four evidence-based reasons for integrating were summarized by McBride and Silverman (1991, p. 286–287): (i) Science and mathematics are closely related systems of thought and are naturally correlated in the physical world; (ii) Science can provide students with concrete examples of abstract mathematical ideas that can improve learning of mathematics concepts;

(iii) Mathematics can enable students to achieve deeper understanding of science concepts by providing ways to quantify and explain science relationships; and (iv) Science activities illustrating mathematics concepts can provide relevancy and motivation for learning mathematics.

Program Context

Although this study focuses on the effect of teacher PD on student achievement in reading, mathematics, and early literacy, the PD is situated within a broader early-childhood science program funded by a Mathematics and Science Partnership (MSP) grant from the National Science Foundation. NURTURES partners a university, local urban public school district, day care centers, and community resources (e.g., science museum, zoo, and botanical gardens) to transform science teaching in local PreK-3 classrooms. The program intervention was designed around the Harvard Complimentary Learning Model (Harvard Family Research Project, 2008) to provide comprehensive educational experiences in science. A complementary learning paradigm makes space for a range of educational experiences that work concurrently as components of a larger process focused on science education.

Schools face intense pressure to improve students' academic achievement, which they must do despite the social problems that students bring to the classroom. Because of this, it would be extremely difficult for schools to alone provide all the supports that children need in their education. Rather, decades of literature suggest that these pressures on schools can be mitigated through partnerships with community agencies and organizations (e.g., Crowson & Boyd, 1993; Heath & McLaughlin, 1987). These connections have been described by Epstein (2011) as a series of overlapping "spheres of influence" that depict the ways in which schools, families, and the community share responsibility for the education of children. In this model, the three players perform different functions, but when they align their goals, the boundaries among them blur, and students receive more coordinated support.

The Harvard Family Research Project (Bouffard, Goss, & Weiss 2008) has grounded their complementary learning framework in this theory. The framework is based on two assertions: first, that both school and non-school contexts makes a critical contribution to students' learning and achievements; and second, that these contexts should create complementary learning opportunities (Weiss, Coffman, Post, Bouffard, & Little, 2005). In such a framework, learning experiences for children are aligned both in school and out of school. This creates a "web of opportunity" for children that breaks down the silos of school, home, and the broader community. In this way, the learning needs of young children are met at school, in after-school programs, in the community, and at home. Rather than each entity operating discretely, a complementary learning paradigm makes space for a range of educational experiences that work concurrently as components of a larger process.

Description of the NURTURES Program

NURTURES includes five primary components: (i) a 2-week Summer Institute for PreK-3 teachers; (ii) academic year PD including monthly professional learning community meetings and one-on-one coaching; (iii) family science activity take-home packs; (iv) family community science events; and (v) public service broadcasts on television that promote family science activities. Figure 1 shows the framework of the program, depicting its intervention components and the intended outcomes.

As seen in Figure 1, two strands of NURTURES address the goal of transforming PreK-3 science teaching: the Summer Institute (SI) and academic year support. Taken together, these two programs constitute at least 92 hours of PD over the course of an academic year, with the bulk of
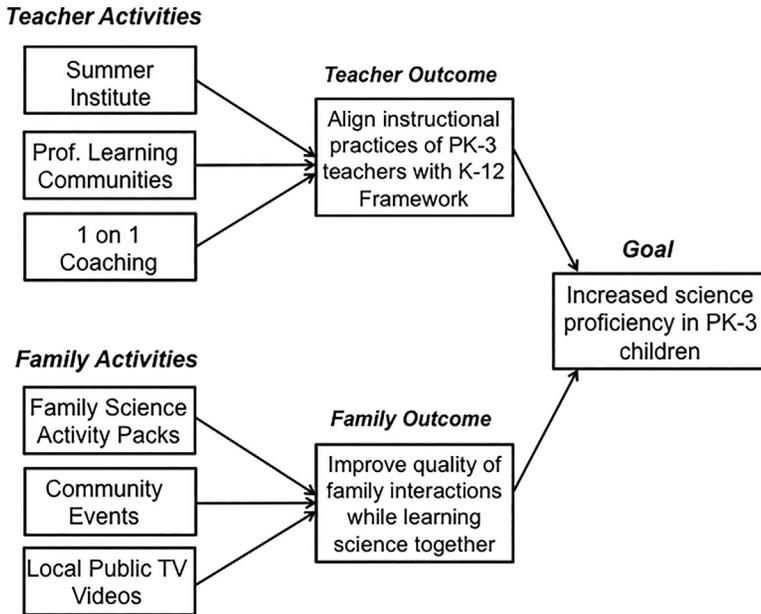
Teacher Activities

Figure 1. Framework of NURTURES.

the hours (80) corresponding to the Summer Institute. As NURTURES focused on four science content sessions and four different grade bands that rotated yearly, teachers could participate in NURTURES for as many as four academic years to experience a full range of science content in their chosen grade band.

Emphasis on Discourse

Especially in early elementary years, when oral language development is a critical component of literacy instruction, student discourse is essential for science learning (NRC, 2007, 2012). The *Framework* (NRC, 2012) emphasizes the importance of students making meaning for themselves, a skill which relies on the development of students' inferential thinking skills. Teachers play an important role in fostering this type of thinking through student discourse. The types of questions or comments that teachers ask during discussions impacts the levels of student thinking that ensue. When adults use open and inferential questions, responses from children demonstrate higher levels of inferential thinking (Danis, Bernard, & Leproux, 2000; Zucker et al., 2010). For this reason, NURTURES professional development for both teachers and families encouraged the use of six "talk moves" to promote discourse (Michaels, Shouse, & Schweinburger, 2008): revoicing, asking someone to restate someone else's reasoning, asking children to apply their own reasoning, prompting children for further participation, asking children to explicate their reasoning, and using wait time. These talk moves encourage children to participate, to collaboratively discuss their thinking, and to reflect, thus increasing their chances to use inferential thinking when engaging in science learning.

NURTURES Summer Institute

The NURTURES SI addresses two main goals: increasing PreK-3 teacher's science content knowledge and understanding of science standards, and aligning their science instruction with the

3D of the *Framework*. A previous study demonstrated the efficacy of the 2014 SI in achieving these goals (Tuttle et al., 2016). These goals were addressed through three main sessions: a content immersion session, a metacognitive session, and a lesson planning session. During these sessions, teachers were provided with a variety of experiences to help them understand and to use the scientific and engineering practices, as well as how to incorporate them into their own teaching. The content immersion and lesson planning portions were split sessions organized by the teachers' grade levels, while the metacognition sections were taught to the full group with all grade levels present.

The content immersion sessions focused on teaching science content through 3D science lessons that were co-taught by a scientist or engineer teamed with an educator. Such a design models the types of science experiences that NURTURES seeks to promote in PreK-3 classrooms. By aligning these lessons to the 3D, teachers gained experience with how to teach disciplinary core ideas through the use of science and engineering practices and an emphasis on crosscutting concepts (See the Supplementary Information online for an example lesson plan). Although the lessons were aimed at adult learners, many lessons could be adapted and taken back to their classrooms to help them teach their grade-level standards. For example, a lesson from the chemistry group focused on density, including how to measure the relevant properties and calculate density from those measurements. While PreK-3 students would not calculate densities, this lesson provided important context for early elementary teachers around the NGSS disciplinary core idea PS1.A (Structure and Properties of Matter). The lesson culminated in a discussion of how best to conduct investigations such as "Sink or Float," which are a staple of early childhood science instruction, to set students up for a later understanding of density. Finally, the lessons extensively incorporated grade-level appropriate expository texts and demonstrated how they might be used to promote science vocabulary and concept learning.

For this portion, the PreK-3 teachers were grouped into grade bands so the content could be contextualized for the grades they teach. The grade bands were PreK-K, K-1, 1–2, and 2–3. Each grade band had a particular content focus for the 2-week SI. In the 2015 SI, PreK-K focused on physical science, K-1 focused on chemical science, 1–2 addressed earth/space science, and 2–3 was geared toward biology.

Metacognitive sessions addressed pedagogical topics appropriate for all teachers. The topics were generally aligned to the *Framework's* science and engineering practices and included facilitation of classroom discourse, developing and using models, asking questions and defining problems, planning and carrying out investigations, using mathematics and computational thinking, vocabulary and expository text, analyzing and interpreting data, and engaging in argument from evidence. These sessions were taught by university faculty members paired with graduate students and NURTURES staff (See the Supplementary Information online for an example lesson plan).

The metacognitive sessions purposefully linked back to information and practices experienced during the content immersion sessions. The goal was to explicitly present the science and engineering practices, provide examples, and foster teacher reflection on the practices. For example, if the morning meta-session focused on classroom discourse, during the hands-on practice sessions the teachers focused on using targeted discourse strategies while working together on an engineering activity. After the activity, the whole group discussed the way the strategies worked and the discourse strategies used during the engineering activity.

Each day, teachers were given time and access to both scientists and their instructional coaches (district teacher leaders who had previously trained with NURTURES) to develop two lessons that they could take back to the classroom and implement during the school year. Teachers were encouraged to incorporate their newly learned skills into their lessons, so that by the end of

the 2 weeks they had written two 3D lessons. During this portion, teachers also had the opportunity to collaborate with their peers to work on their lessons and generate ideas to address their standards (See the Supplementary Information online for the 3D lesson plan template).

Academic Year Support

During the academic year, classroom teachers participated in six 2-hour Professional Learning Communities (PLCs), which were facilitated by NURTURES staff. Topics of study for the PLCs built on ideas introduced during the SI or focus on teacher interests or identified challenges (e.g., classroom management or assessment). PLCs were generally structured such that NURTURES staff set the agenda for sessions early in the year, and teachers set the agenda for sessions later in the year. Topics varied across PLC sessions, as each group chose the questions it wanted to address. However, topics that arose frequently included student discourse—how to ask questions that would promote inferential thinking, and how to encourage and teach student-to-student discourse, along with discussions that focused on how to foster more data analysis and interpretation following a classroom investigation. Since the PLCs were organized by grade level, the sessions provided a timely opportunity throughout the school year for teachers to *confer* about the science standards they were teaching during a particular quarter.

Teachers were also given the option of *one-to-one coaching* throughout the academic year, and approximately 33% of teachers choose individualized coaching. This option was available to all teachers in 2013–2014 due to the smaller number of participants (40) that year but was limited due to staff capacity in 2014–2015 and 2015–2016, when there were approximately 150 participants each year. Six coaching cycles occurred each academic year. The coaching cycle consisted of five steps: (i) teachers sent their coach a lesson plan to be reviewed; (ii) the coach makes comments and suggestions prior to the lesson; (iii) the coach observes the teacher's lesson; (iv) the teacher and coach discuss any pedagogical practices needing attention; and (v) teacher and coach work together to develop goals for the next lesson. As with the PLCs, areas of focus for teacher–coach pairs varied significantly. One major area of focus, as with the PLCs, was student discourse, including how to ask questions that would promote inferential thinking, and how to encourage and teach student-to-student discourse. In sum, how to build and enhance higher levels of student thinking and talking was a consistent theme across many, if not most, of the individual coaching sessions.

Family Engagement

To engage families of young children in science learning, NURTURES uses two strategies: one is classroom-based (Family Packs), and the other is community-based (family community science events).

Family Packs. Twenty family activity packs for five different grade levels (PreK, Kindergarten, 1st, 2nd, and 3rd) were sent home by teachers to make home-school connections in science and encourage family science investigations and discourse. Each Family Pack contained interesting science activities aligned with the SEPs in the 3D in the *Framework* (NRC, 2012) and various early learning standards (e.g., NAEYC, 2016; ODE, 2012). The Family Packs were designed to help parents know what to do, have access to background knowledge, and promote high quality discourse (Reinhart et al., 2016). Each pack included a newsletter with the directions for the investigation, necessary materials for the activity, and a journal sheet for children to record data or visually represent understanding. The packs have been translated to Spanish for families where English is the second language.

Family Community Science Events. Family community science events were offered to give families opportunities to engage in informal science activities in the community (e.g., a park, zoo, science center, public library, farm). A wide range of activities such as engineering challenges, simulations, observations, and demonstrations, were geared for families of young children and were designed to foster adult–child interaction around a variety of science topics. An "event guide" helped adults to facilitate their child's experience, providing needed scaffolds for each activity. These guides included step-by-step directions for adults, suggestions for language to use, questions to ask, and spaces to record children's responses. Additional information and detailed ideas for engaging children were on the back.

Methodology

This study used a quasi-experimental, between-group design to investigate whether NURTURES affected student learning outcomes. To that end, the study asks, *what effect does teachers' participation in NURTURES have on their students' achievement in early literacy, reading, and mathematics*?

Participants

Control and treatment students were drawn from students at the 41 elementary schools in a large urban school district in the Midwest with a high degree of racial diversity and 64.8% of students receiving free and reduced lunch. Since this study examined students in three different assessments (early literacy, mathematics, and reading), which were administered by the district in different grades, the three sub-studies have different participant populations. Students were drawn from grades K-2 (early literacy), grades 2–4 (mathematics), and grades 1–4 (reading) based on the district's timetable for assessment. Differences are seen among the sample sizes due to selecting students from different grades for each assessment and because NURTURES teachers are not randomly distributed among grades.

Teachers were recruited to NURTURES in three stages. First, a targeted recruitment was conducted for eight teacher leaders, who completed the SI during 2012. Teacher leaders went through a competitive application process before acceptance. They then assisted NURTURES staff with lesson planning during subsequent years. For the 2013 SI, which had 40 participants, NURTURES was open to any PK-3 teacher from the teacher leaders' schools. In 2014 and 2015, NURTURES was open to any PK-3 teacher from the district.

Treatment participants consisted of students who had had at least one NURTURES teacher during the 2013–2014, 2014–2015, or 2015–2016 academic years. Teachers' participation in NURTURES could have occurred in any or multiple of those academic years. Participants consisted of 2,899 students for the early literacy study, 2002 students for the mathematics study, and 1,810 students for the reading study. Control students consisted of 2,515 students for the early literacy study, 3,028 students for the mathematics study, and 2,448 students for the reading study, who had never had a NURTURES teacher within the same time frame.

Data

Data consisted of raw and scale scores from the STAR Early Literacy, Mathematics, and Reading assessments. These nationally normed assessments are grounded in research and have been reviewed as reliable and valid by several independent groups (Renaissance Learning, 2014). These assessments were chosen due to their availability as validated assessments for the grade range of interest to NURTURES, and because their use as a formative assessment by the district meant that students were assessed multiple times throughout the academic year.

STAR Early Literacy is a 27-item adaptive assessment consisting of operational items aligned to early literacy skills derived from state standards, the Common Core standards, and current research. The content is organized into 3 broad domains (Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning, and Numbers and Operations), and 10 sub-domains (Alphabetic Principle, Concept of Word, Visual Discrimination, Phonemic Awareness, Phonics, Structural Analysis, Vocabulary, Sentence-Level Comprehension, Paragraph-Level Comprehension, and Early Numeracy).

STAR Mathematics is a 24-item assessment focusing on problem solving, reasoning and proof, communication, representation, connections, adaptive reasoning, strategic competence, conceptual understanding, procedural fluency, and productive disposition. STAR Reading is a fixed-length adaptive test with 25 items in grades K-2 and 20 items in grade 3. This assessment focuses on vocabulary in context and reading comprehension.

These assessments provided a natural pre-/post-framework for this study, as our partner district administers them in both fall and spring. Data were collected from the district for academic year 2015–2016, which included three measurement occasions: Fall 2015, Winter 2015, and Spring 2016. For the grades of interest to this study, the district administers STAR Early Literacy in grades K-2, STAR Mathematics in grades 2–4, and STAR Reading in grades 2–4. In addition, K-1 students who achieve a threshold on the STAR Early Literacy assessment are given the STAR Reading assessment before grade 2. Based on this schedule, the data selected for analysis were drawn from grades K-2 (STAR Early Literacy), grades 2–4 (STAR Mathematics), and grades 1–4 (STAR Reading). In addition, data from STAR Early Literacy assessments given to kinder-garteners in Fall 2013 and Fall 2014 were examined to establish baseline equivalence (see below).

Baseline Equivalence

Baseline equivalence was established by examining the fall scores for the STAR Early Literacy assessment for kindergarteners in the study for 2013–2014, 2014–2015, and 2015–2016. This was necessary due to the nature of the NURTURES intervention coupled with the district's schedule for assessment. First, the partner district does not offer STAR Mathematics in grades K-1, and only offers STAR Reading in grades K-1 to students who are reading already. Second, the intervention examined in this study occurred for teachers in grades K-3. Taken together, this means that, in the 2015–2016 year of interest to this study, students in grades 2–4 could have had an intervention teacher in 2013–2014, 2014–2015, or 2015–2016. This confounds a comparison of the STAR Reading and Mathematics performances of the control and intervention students in grades 2–4 in Fall 2015. We, therefore, examined the equivalence of students at the beginning of their academic careers by examining Fall scores for Kindergarteners from 2013 to 2014, 2014 to 2015, and 2015 to 2016 on the STAR Early Literacy assessment. This assessment contains items foundational for later reading and mathematics skills, particularly in the domains of Comprehen-sion & Skills and Numeracy & Operations. Using STAR Early Literacy to assess students' skill levels in the relevant domains avoids this potential confounding factor.

A two-level hierarchical model was used to assess the equivalency between the treatment and control cohorts; three separate analyses were performed for the 3 respective years. The first level equation that predicted individual student's mean achievement included the intercept value and the participant's *gender*, *ethnicity*, *intervention type* variables and a random error component. To capture the effects of individual schools, a second-level, unconditional equation was added. The intercept and *intervention type* coefficients were considered random, and the effects of *gender* and *ethnicity* were considered fixed at the school level. The results of the three separate analyses demonstrating the equivalency of K-students in each year are summarized in Tables S1–S3 in the Supplementary Information. The results for the treatment type coefficients for all three years

indicated no statistically significant difference between the groups: $t(40) = -6.66$, $p = 0.242$ for fall 2013; $t(40) = 3.32$, $p = 0.359$ for fall 2014; and $t(40) = 0.87$, $p = 0.777$ for Fall 2015 data. The weighted average absolute value effect size for intervention (Hedges' g) was 0.047, which is considered to be a negligible effect size, so no statistical correction for baseline was used during subsequent data analyses.

Data Analysis

Hierarchical Linear Models. Objectively assessing student growth through large-scale data, such as students' standardized test results in particular school districts that includes dozens of schools and thousands of students, presents several methodological challenges. The most significant difficulty in modeling student achievement relates to the hierarchical nature of data. Students are not randomly distributed; rather, individual students are "nested" within classrooms within particular schools. The modeling challenges a researcher faces when dealing with these complex data structures include accounting for a number of factors, such as, (i) determining student performance while also accounting for student-level factors such as gender and/or minority status and (ii) accounting for group effects that influence all the children in the classroom, such as the teacher's level of experience or the school's overall level of poverty. If the researcher uses conventional methods and computes separate regression equations for each school and specific subgroups, the result is a proliferation of separate regression equations, which do not permit global analysis of the data. On the other hand, if the researcher aggregates data within or across schools, the data will not reflect the nuanced impact on an individual child who is taught by a specific teacher within a particular school. The use of hierarchical linear models resolves this problem of nested data by modeling equations at each level of the data structure and then including factors in the overall equation such that lower level data analyses loop into subsequent layers of data analysis.

This Study. The hierarchical model adopted in this study is a three-level hierarchical model, as implemented by HLM for Windows, v. 7.01, where the students' Rasch-model scaled STAR scores for the Mathematics, Reading, and Early Literacy assessments serve as the outcome variables. Although a multivariate approach to the dependent variables is possible, the present study focused on the analysis of one outcome variable at a time. Therefore, the first-level of data consists of repeated observations of the assessment data in one domain (a level-one unit) nested within a specific student (a level-two unit). Students in turn are nested within schools (a level-three unit).

At the first level equation, the individual student mean achievement was predicted from one time-variant variable: grand-mean centered *testing occasion* (levels: 0 = Fall 2015, 1 = Winter 2015, and 2 = Spring 2016). The first-level equation included student's intercept (mean value of student achievement) and his/her slope or individual growth trajectory over the measurement occasions, plus a random error interpreted as a residual temporal variation. At the second-level, the estimated coefficients (intercepts and slopes) from the first-level equations became the solutions to two equations, one that modeled student's mean achievement or π_{0jk} and another one that modeled student average learning rate or π_{1jk}. Both second level equations included time-invariant student-level variables: grand-mean centered *current grade* (2, 3, and 4 for mathematics; 1, 2, and 3 for reading; K, 1, and 2 for early literacy); *gender* (levels: 0 = female and 1 = male); *minority status* (levels: 0 = minority or and 1 = non-minority or White); and *intervention* or the number of teachers the student had up to the point of measurement (levels: 0, 1, 2, where 2 represented having

two or more intervention teachers). The *current grade* variable was considered a time-invariant because the assessment data utilized the latest, 2015–2016 academic year data. The effects of schools were modeled with the third-level equations. The third-level equations were unconditional or did not include school-context variables.

The equations below depict the specification of the model at each of the three levels. The variance components were specified as random at a student-level. With respect to the school-level, the effects of *gender* and *minority status* are assumed to be invariant between schools, while the effects of the *current grade* and *intervention* are assumed to be random.

Level 1 Model

$$SCALEDSC_{ijk} = \pi_{0jk} + \pi_{1jk}{}^{*}(OCCASION_{ijk}) + e_{ijk}$$

Level 2 Model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}{}^{*}(CURRENTG_{jk}) + \beta_{02k}{}^{*}(GENDER\_M_{jk}) + \beta_{03k}{}^{*}(MINORITY_{jk})$$
$$+ \beta_{04k}{}^{*}(I\_T_{jk}) + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + \beta_{11k}{}^{*}(CURRENTG_{jk}) + \beta_{12k}{}^{*}(GENDER\_M_{jk}) + \beta_{13k}{}^{*}(MINORITY_{jk})$$
$$+ \beta_{14k}{}^{*}(I\_T_{jk}) + r_{1jk}$$

Level 3 Model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{01k} = \gamma_{010} + u_{01k}$$
$$\beta_{02k} = \gamma_{020}$$
$$\beta_{03k} = \gamma_{030}$$
$$\beta_{04k} = \gamma_{040} + u_{04k}$$
$$\beta_{10k} = \gamma_{100} + u_{10k}$$
$$\beta_{11k} = \gamma_{110} + u_{11k}$$
$$\beta_{12k} = \gamma_{120}$$
$$\beta_{13k} = \gamma_{130}$$
$$\beta_{14k} = \gamma_{140} + u_{14k}$$

Mixed Model

$$SCALEDSC_{ijk} =$$
$$\gamma_{000} + \gamma_{010}{}^{*}CURRENTG_{jk} + \gamma_{020}{}^{*}GENDER\_M_{jk} + \gamma_{030}{}^{*}MINORITY_{jk} + \gamma_{040}{}^{*}$$
$$I\_T_{jk} + \gamma_{100}{}^{*}OCCASION_{ijk} + \gamma_{110}{}^{*}OCCASION_{ijk}{}^{*}CURRENTG_{jk} + \gamma_{120}$$
$$^{*}OCCASION_{ijk}{}^{*}GENDER\_M_{jk} + \gamma_{130}{}^{*}OCCASION_{ijk}{}^{*}MINORITY_{jk} + \gamma_{140}{}^{*}$$
$$OCCASION_{ijk}{}^{*}I\_T_{jk} + r_{0jk} + r_{1jk}{}^{*}OCCASION_{ijk} + u_{00k} + u_{01k}{}^{*}CURRENTG_{jk} +$$
$$u_{04k}{}^{*}I\_T_{jk} + u_{10k}{}^{*}OCCASION_{ijk} + u_{11k}{}^{*}OCCASION_{ijk}{}^{*}CURRENTG_{jk} + u_{14k}{}^{*}$$
$$OCCASION_{ijk}{}^{*}I\_T_{jk} + e_{ijk}$$

Results

This study examines the impact of a NURTURES teacher on student achievement in early literacy, reading, and mathematics, assessed in the context of a repeated measurement. Students' changes in their STAR standardized scores were examined as a function of student-level

characteristics, grade level, gender, minority status, and the number of intervention teachers (0, 1, or 2 or more). Three school year-end assessments were modeled to be nested within each student, and each student's mean and growth components in respective areas were assessed in the context of a school the student belonged to. The HLM, therefore, partitioned students' variance in mean and growth status due to within- and between-school effects. As a result, a measure of participant's performance could be obtained, regardless of the school the student happened to be in. The model accomplishes the analysis by modeling a unique amount of variance associated with each school and explicitly incorporates school-level variance when estimating individual student performance.

STAR Early Literacy

Fitting the STAR Early Literacy data to the hierarchical linear model resulted in a student mean achievement (expressed as a γ_{000}, the third-level equation intercept coefficient) of 649.91 (see Table 1). This coefficient represented an average, predicted Winter 2015 score for a minority, female 1st grade student who had never had a NURTURES teacher. This model-predicted outcome was affected in a statistically significant way by several student-level demographic variables. As expected, students' scale scores increased by 101.43 with an increase in *current grade*, expressed as the γ_{010} coefficient (i.e., moving from grades 1 to 2) when controlling for the effects of *gender*, *minority status*, and *intervention*. The effect of *gender* (the

Table 1

Summary of three-level exploratory model for STAR early literacy achievement

Fixed Effect	B	SE B	t-Ratio	df	p
Model for average status, π_0					
Model for mean-status of 1st grade minority female who did not have intervention teacher between, β_{00}					
Average mean status, γ_{000}	649.91	7.68	84.64	41	<0.001
Model for current grade, β_{01}					
Current grade, γ_{010}	101.43	2.19	46.32	41	<0.001
Model for gender, β_{02}					
Gender, γ_{020}	−14.69	2.00	−7.35	5,854	<0.001
Model for minority status, β_{03}					
Minority status, γ_{030}	15.37	2.18	7.07	5,854	<0.001
Model for cumulative intervention, β_{04}					
Cumulative intervention, γ_{040}	8.59	3.56	2.41	41	0.020
Model for learning rates, π_1					
Model for learning rates of 1st grade minority female who did not have intervention teacher between, β_{10}					
Average learning rate, γ_{100}	68.51	1.85	37.07	41	<0.001
Model for current grade, β_{11}					
Current grade, γ_{110}	−18.95	2.03	−9.32	41	<0.001
Model for gender, β_{12}					
Gender, γ_{120}	0.88	1.34	0.66	5,854	0.512
Model for minority status, β_{13}					
Minority status, γ_{130}	0.19	1.41	0.14	5,854	0.892
Model for cumulative intervention, β_{14}					
Cumulative intervention, γ_{140}	−2.65	1.47	−1.80	41	0.079

γ_{020} coefficient) on mean achievement status was statistically significant, with female students outscoring male students by an average of 14.69 units. Also, a statistically significant effect for *minority status* (the γ_{030} coefficient) on mean achievement was observed, with non-minority students scoring, on average, an additional 15.37 units in comparison to minority students. This final effect, however, has to be interpreted cautiously in the absence of student's socio-economic status information.

Most importantly, the *intervention* variable had a statistically significant impact on students' scores (see the γ_{040} coefficient). Adding an additional NURTURES teacher to a student's academic history was associated with an average increase of 8.59 units in mean student achievement, controlling for the effects of the *current grade*, *gender*, and *minority status* variables. This effect size (Hedges' *g*) was 0.066, which is to be interpreted as a treatment group having, on average, 0.066 higher scores in standard deviation units as compared to the scores of the control cohort and is to be interpreted a small effect size.

This model also provided information about the associated changes in student mean achievement score from one testing occasion to another, or a learning rate, expressed as the γ_{100} coefficient, which is also included in Table 1. The learning rate for a minority, 1st grade female student who had never had a NURTURES teacher was 68.51 units, meaning that this hypothetical student's score rose 68.51 units from fall to spring. No student-level variables, with the exception of *current grade* (see the γ_{110} coefficient) had a statistically significant effect on the learning rate over this relatively short assessment time. On average, students in lower grades experienced 18.95 units faster learning than students in higher grades over testing occasions (see the γ_{140} coefficient), when controlling for the effects of *gender*, *minority status*, and *intervention*. As the reliability of the estimate of the mean learning rate was low (see below), these results should be interpreted cautiously.

Reliability coefficients for this model, represented as the absolute values of variance, τ_π and τ_β, which measure the amount of "signal" in the data, are summarized in Table S1. Bryk and Raudenbush (1992) consider values above 0.2 as acceptable. All of the reliability values are in the acceptable range, with the exception for the estimate of the mean learning rate, π_1, which is 0.099. Therefore, the model of the mean learning rate should be interpreted with caution.

Finally, when examining the variance of the conditional model (see Table S2), 64.09% of the variance on mean status and 61.16% of the variance in mean learning curve were explained by student-level variables. See Table S3 for an explanation of the variance components of an unconditional model.

STAR Mathematics

Fitting the STAR Mathematics data with the HLM resulted in a student mean achievement (expressed as a γ_{000}, the third-level equation intercept coefficient) of 494.34 (see Table 2). This coefficient represented an average, predicted Winter 2015 score for a minority, female 3rd grade student who had never had a NURTURES teacher. Three of the four student-level variables had a statistically significant effect on the mean measure. The effect of *gender* on a student mean achievement status was not statistically significant (see the γ_{020} coefficient). However, students' scale scores increased by 85.25 units with an increase in *current grade* (i.e., moving from grade three to grade four) when controlling for the effects of *gender*, *minority status*, and *intervention* (see the γ_{010} coefficient). A statistically significant effect for *minority status* on mean achievement was observed, with non-minority students scoring, on average, an additional 21.21 units in comparison to minority students (see the γ_{030} coefficient). This effect, again, should be interpreted cautiously in the absence of student's socio-economic status information.

Table 2

Summary of three-level exploratory model for STAR mathematics achievement

Fixed Effect	B	SE B	t-Ratio	df	p
Model for average status, π_0					
Model for mean-status of 3rd grade minority female with no intervention β_{00}					
Average mean status, γ_{000}	494.35	5.73	86.31	40	<0.001
Model for current grade, β_{01}					
Current grade, γ_{010}	85.25	2.09	40.74	40	<0.001
Model for gender, β_{02}					
Gender, γ_{020}	2.76	2.11	1.31	5,537	0.190
Model for minority status, β_{03}					
Minority status, γ_{030}	21.21	2.33	9.1	5,537	<0.001
Model for cumulative intervention, β_{04}					
Cumulative intervention, γ_{040}	16.99	2.56	6.63	40	<0.001
Model for learning rates, π_1					
Model for learning rates of 3rd grade minority female with no intervention, β_{10}					
Average learning rate, γ_{100}	47.76	1.71	27.94	40	<0.001
Model for current grade, β_{11}					
Current grade, γ_{110}	−6.76	1.58	−4.28	40	<0.001
Model for gender, β_{12}					
Gender, γ_{120}	2.28	1.06	2.16	5,537	0.031
Model for minority status, β_{13}					
Minority status, γ_{130}	0.93	1.16	0.81	5,537	0.418
Model for cumulative intervention, β_{14}					
Cumulative intervention, γ_{140}	0.78	1.57	−0.5	40	0.622

Most importantly, the intervention variable has a statistically significant impact on students' mean achievement on the STAR Mathematics assessment (see the γ_{040} coefficient). An average increase of 16.98 units was associated with adding an additional NURTURES teacher to a student's academic history, controlling for the effects of the *current grade*, *gender*, and *minority status* variables. This effect size (Hedges' *g*) was calculated as 0.140.

Analogously, with respect to the assessment of a student's learning rate, the average slope coefficient for a minority, 3rd grade female student who had never had a NURTURES teacher was 47.76 units (see the γ_{100} coefficient in Table S4). No student-level variables, with the exception of *current grade*, had a statistically significant effect on the learning rate over this relatively short assessment time. On average, students in higher grades increase their scores at 6.76 units slower than students in lower grades, when controlling for the effects of *gender*, *minority status*, and *intervention* (see the γ_{110} coefficient).

Reliability coefficients for this model are summarized in Table S4. All of the reliability values are in the acceptable range, and the reliability coefficients from an exploratory model were relatively unaffected when compared to the reliability coefficients estimated for an unconditional model. Finally, when examining the variance of the conditional model (see Table S5), 48.40% of the variance on mean status and 14.22% of the variance in mean learning curve were explained by student-level variables. See Table S6 for an explanation of the variance components of an unconditional model.

Table 3

Summary of three-level exploratory model for STAR reading achievement

Fixed Effect	B	SE B	t-Ratio	df	p
Model for average status, π_0					
Model for mean-status of minority					
female who did not have intervention					
teacher between grades 2 and 3, β_{00}					
Average mean status, γ_{000}	301.71	8.68	34.74	40	<0.001
Model for current grade, β_{01}					
Current grade, γ_{010}	79.56	3.27	24.35	40	<0.001
Model for gender, β_{02}					
Gender, γ_{020}	−14.19	3.53	−4.02	4,952	<0.001
Model for minority status, β_{03}					
Minority status, γ_{030}	42.15	3.91	10.78	4,952	<0.001
Model for cumulative intervention, β_{04}					
Cumulative intervention, γ_{040}	41.38	4.17	9.93	40	<0.001
Model for learning rates, π_1					
Model for learning rates of minority					
female who did not have intervention					
teacher between grades 2 and 3, β_{00}					
Average learning rate, γ_{100}	53.55	2.07	25.86	40	<0.001
Model for current grade, β_{11}					
Current grade, γ_{110}	−3.65	1.47	−2.49	40	0.017
Model for gender, β_{12}					
Gender, γ_{120}	1.65	1.69	0.98	4,952	0.328
Model for minority status, β_{13}					
Minority status, γ_{130}	4.92	1.81	2.71	4,952	0.007
Model for cumulative intervention, β_{14}					
Cumulative intervention, γ_{140}	−0.02	1.94	−0.01	40	0.992

STAR Reading

    Fitting the STAR Reading data to the hierarchical linear model resulted in a student mean achievement (expressed as a γ_{000}, the third-level equation intercept coefficient) of 301.71 (see Table 3). This coefficient represented an average, predicted Winter 2015 score for a minority, female student between grades 2 and 3 who had never had a NURTURES teacher. The examination of the student-level variables included in the model demonstrated statistically significant effects for all of the second-level variables. Students' scale scores increased by 79.56 with an increase in *current grade* (i.e., moving from grade three to grade four) when controlling for the effects of *gender*, *minority status*, and *intervention* (see the γ_{010} coefficient). A statistically significant effect for *gender* (see the γ_{020} coefficient) on mean achievement was observed with female students gaining an additional 14.19 units in comparison to male students. A statistically significant effect for *minority status* on mean achievement was present, with non-minority students scoring, on average, an additional 42.15 units in comparison to minority students (see the γ_{030} coefficient). Again, this effect should be interpreted cautiously in the absence of student's socio-economic status information.

    Most importantly, the intervention variable has a statistically significant impact on students' mean achievement on the STAR Reading assessment (see the γ_{040} coefficient). An average increase of 41.38 units was calculated as a function of adding a NURTURES teacher to a student's academic history, controlling for the effects of the *current grade*, *gender*, and *minority status*

variables. This effect size (Hedges' g) was calculated as 0.250, a level considered substantively important by the What Works Clearinghouse (What Works Clearinghouse, 2013).

As with the STAR Early Literacy and Mathematics models, this model also provided information about the increase in score from one testing occasion to another, or learning rate. The learning rate for a minority female student who had never had a NURTURES teacher, see the γ_{100} coefficient in Table S7, was 53.55 units. Most student-level variables had small, statistically significant effects on the learning rate. The effect of *current grade* (see the γ_{110} coefficient) on the learning rate was statistically significant, with students in higher grades learning at 3.65 units slower than students in lower grades. The growth differential for *minority status* (see the γ_{130} coefficient) was also statistically significantly different, with non-minority students making 4.92 unit gains more than non-minority students from one testing occasion to another. Also, the effect of *gender* (see the γ_{120} coefficient), controlling for the effects of *current grade*, *minority status*, and *intervention*, was statistically significant, with males outgrowing females by an average of 2.28 units between assessment times.

Reliability coefficients for this model are summarized in Table S7. All of the reliability values are in the acceptable range, and the reliability coefficients from an exploratory model were relatively unaffected when compared to the reliability coefficients estimated for an unconditional model. Finally, when examining the variance of the conditional model (see Table S8), 31.52% of the variance on mean status and 4.75% of the variance in mean learning curve were explained by student-level variables. See Table S9 for an explanation of the variance components of an unconditional model.

Discussion

This study provided evidence for the efficacy of NURTURES in affecting student outcomes in early literacy, reading, and mathematics when student level variables, namely gender, ethnicity, and grade level were considered and the school context or between-schools variation properly accounted for. Having a NURTURES teacher in the student's academic life prior or during 2015–2016 school year was associated with net gains of 8.6 points to a student's STAR Early Literacy spring score, 17.0 points to a student's STAR Mathematics spring score, and 41.4 points to a student's STAR Reading spring score compared to students who had never had a NURTURES teacher. The 41.4 points in STAR Reading translated to an effect size of 0.25, a level considered substantively important by the What Works Clearinghouse evidence standards (What Works Clearinghouse, 2013).

Although this study represents an important step toward understanding the effect of a *Framework*-aligned science PD program on student academic achievement, it leaves open the question of causality, which will require a randomized control trial to test. Furthermore, within NURTURES, teachers have significant freedom to design *Framework*-aligned lessons, and this study does not distinguish which aspects of the teachers' instruction contribute to the achievement gains associated with NURTURES teachers. Out-of-school components may have contributed; all NURTURES teachers were given *Framework*-aligned family packs to send home with students, and all of their students were invited to the six annual family community science events. Evidence from other studies suggests possible contributors from teachers' in-school practices. The experiences of NURTURES coaches and PLC facilitators indicated that teachers, year after year, wanted to focus much of their academic year PD on enhancing student discourse during science instruction. Our previous results (Tuttle et al., 2016) demonstrated that a sample of eleven teachers did incorporate discourse strategies, expository text, and vocabulary into a majority of their science lesson plans and science instruction following the 2014 NURTURES SI. Furthermore, they also incorporated mathematical and computational thinking into their lesson plans and

instruction, albeit to a lesser extent. Additional factors could include increased time spent on science instruction or the integrative science instruction envisioned in the *Framework*. Further research will need to be conducted to identify which factors or combination of factors contributed most to the gains measured here.

The different effect sizes measured in this study raise the question of analysis of the domains tested within each STAR assessment also illuminates our interpretation of the student academic changes show in this study. For example, it is unsurprising that the effect size on STAR Early Literacy scores is smaller than for STAR Mathematics or STAR Reading given that the STAR Early Literacy assessment focuses on domains foundational for later reading and math skills. While some early literacy sub-domains integrate well with science instruction (e.g., Vocabulary), other assessed early literacy domains are less frequently integrated into science instruction (e.g., Phonics, Concept of Word, and Phonemic Awareness). The variation between student literacy growth versus mathematic growth documented in the current study may be influenced by the amount of exposure to language-based versus math-based instruction presented by NURTURES teachers. Teachers may have replicated in their classrooms this pattern of a heavier emphasis on language-based instruction in science, with a lesser emphasis on bringing math-based instruction into science.

Overall, this study demonstrated that a *Framework*-aligned PD for early elementary educators can potentially lead to gains in student achievement in literacy, reading, and mathematics. Our work, therefore, supports the idea that *Framework*-aligned science should not be considered an "extra" classroom box to check off but rather a model for contextualizing the teaching of literacy, reading, and mathematics.

Preparing Teachers for Early Childhood Science

These results have implications for designers of science PD aimed at in-service early elementary educators. First, aligning science PD with the 3D *Framework* (NRC, 2012) may contribute to the gains seen in this study, as the science and engineering practices previously observed from NURTURES teachers (Tuttle et al., 2016) align with the student gains measured in this study. Therefore, aligning PD to the NGSS and *Framework* (NRC, 2012) may represent a way to help teachers effectively teach science in their classrooms, and also to help them meet student learning goals for literacy, reading, and mathematics in their districts. This alignment may reinforce best practices in literacy, reading, and mathematics in a science context. The NURTURES focus on literacy and reading included the use of expository text in science lessons, teaching vocabulary, and promoting inferential thinking and scientific reasoning with student discourse. Mathematics integration was demonstrated via direct instruction on using mathematics and computational thinking and a strong focus on data analysis, including creating graphs and charts and interpreting their meaning, during science investigations.

In addition, these results have implications for pre-service teacher preparation programs. In general, early elementary teachers receive limited domain-specific instruction such as science, with the exception of literacy during their teacher preparation programs (Isenberg, 2000; Lobman, McLaughlin, & Ryan, 2005). For that reason, among others, early childhood and early elementary teachers often feel underprepared to teach science (Banilower et al., 2013). The need to better prepare these teachers is becoming expert consensus; a recent report from the National Research Council notes that, "Content and methods courses in higher education, as well as other professional learning activities, need to enhance the competencies of educators of children from birth through age 8 in all aspects of science learning trajectories: science goals and content, developmental progressions for a variety of science topics, and instructional tasks and strategies" (Institute of Medicine and National Research Council, 2015, p. 271).

Implications for Policy

Considering the recent emphasis on science instruction for early childhood classrooms, such as the April 2016 White House event in support of several public and private initiatives focusing on STEM for young children (Samuels, 2016), it is worthwhile to consider the implications of this work for science policy. Achievement gaps in literacy and numeracy in early childhood, which have repeatedly been shown to predict later reading and mathematics achievement gaps (e.g., Chatterji, 2006; Downey, von Hippel, & Broh, 2004; Jordan, Kaplan, Ramineni, & Locuniak, 2009), receive significant attention through programs such as Head Start and Early Reading First. However, the science achievement gap receives less attention (Tate, Jones, Thorne-Wallington, & Hogrebe, 2012). This is significant because recent work demonstrates that the science achievement gap begins in kindergarten and persists at least to eighth grade (Morgan, Farkas, Hillemeier, & Maczuga, 2016). Furthermore, elementary instructional time for science, which can increase science achievement, has been dropping in the United States (Blank, 2013).

This study leaves open the question of the impact of NURTURES on the academic achievement of preschoolers, which future research should address. However, this study does suggest that achievement gaps in reading, and mathematics can be addressed in part by providing *Framework*-aligned science instruction in early elementary classrooms. Including science instruction in early childhood and early elementary classrooms provides opportunities to increase science achievement (Blank, 2013); our work suggests that aligning that science instruction with the *Framework* can improve students' achievement in early literacy, reading, and mathematics. We note that the gains measured for students in mathematics and reading in this study were comparable to the gaps measured for minority and non-minority students. Varelas et al. (2014) also demonstrated that students of color, when given access to quality science instruction that accounts for the knowledge that they bring to the classroom, demonstrate the kind of scientific thinking advocated by reform proposals. Future research must be conducted to identify the most important determinants of teachers' *Framework*-aligned instruction for the gains measured here, and policymakers could use that information to support NGSS implementation and the adoption of *Framework*-aligned science curricula in early childhood and early elementary classrooms as a means for reducing achievement gaps in science, reading, and mathematics.

References

Banilower, E. R., Smith, S., Weiss, I. R., Malzahn, K. J., Campbell, K. H., & Weis, A. H. (2013). Report of the 2012 national survey of science and mathematics education. Chapel Hill, NC: Horizon Research, Inc.

Blank, R. K. (2013). Science instructional time is declining in elementary schools: What are the implications for student achievement and closing the gap?. Science Education, 97, 830–847. https://doi.org/ 10.1007/ s11191-012-9537-6

Bouffard, S. M., Goss, C. B., & Weiss, H. B. (2008). Complementary learning: Emerging strategies, evolving ideas. Cambridge, MA: Harvard Family Research Project.

Breslich, E. R. (1936). Integration of secondary school mathematics and science. School Science and Mathematics, 36(1), 58–67.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage Publications.

Capps, D. K., Crawford, B. A., & Constas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. Journal of Science Teacher Education, 23(3), 291–318.

Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. G. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. Journal of Research in Science Teaching, 49, 631–658.

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth grade slump. American Educator, 27(1), 14–16.

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the early childhood longitudinal study (ECLS) kindergarten to first grade sample. Journal of Educational Psychology, 98, 489–507. https://doi.org/10.1037/0022-0663.98.3.489

Crowson, R. L., & Boyd, W. L. (1993). Coordinated services for children: Designing arks for storms and seas unknown. American Journal of Education, 101(2), 140–179.

Czerniak, C. M., & Johnson, C. C. (2014). Interdisciplinary science and STEM teaching. In N. G. Lederman, & S. K. Abell (Eds.), Handbook of research on science education (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Danis, A., Bernard, J., & Leproux, C. (2000). Shared picture-book reading: A sequential analysis of adult-child verbal interactions. British Journal of Developmental Psychology, 18, 369–388.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. American Sociological Review, 69, 613–635.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. C., Crawford, G. M., . . . Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. Early Childhood Research Quarterly, 25, 177–193. https://doi.org/10.1016/j.ecresq.2009.10.003

Epstein, J. L. (2011). School, family, and community partnerships: Preparing educators and improving schools. Boulder, CO: Westview Press.

Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. International Electronic Journal of Elementary Education, 4(1), 5.

Glynn, S., & Muth, K. D. (1994). Reading and writing to learn science: Achieving scientific literacy. Journal of Research in Science Teaching, 31(9), 1057–1073.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: A programmatic research agenda to improve science readiness. Early Education and Development, 20, 238–264. https://doi.org/10.1080/10409280802595441

Guthrie, J. T., & Ozgungor, S. (2002). Instructional contexts for reading engagement. In C. C. Block, & M. Pressley (Eds.), Comprehension instruction: Research-based best practices (pp. 275–288). New York: The Guilford Press.

Hapgood, S., & Palincsar, A. M. (2007). Where literacy and science intersect. Educational Leadership, 64(4), 56–60.

Heath, S. B., & McLaughlin, M. W. (1987). A child resource policy: Moving beyond dependence on school and family. Phi Delta Kappan, 68(8), 576–580.

Holliday, W. G., Yore, L. D., & Alvermann, D. E. (1994). The reading-science learning-writing connection: Breakthroughs, barriers, and promises. Journal of Research in Science Teaching, 31(9), 877–893.

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

Institute of Medicine and National Research Council. (2015). Transforming the workforce for children birth through age 8: A unifying foundation. Washington, DC: The National Academies Press.

Isenberg, J. P. (2000). The state of the art in early childhood professional preparation. New teachers for a new century: The future of early childhood professional preparation (pp 15–58). Washington, DC: National Institute on Early Childhood Development and Education.

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. Phi Delta Kappan, 81, 199–203.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. Developmental Psychology, 45, 850–867. https://doi.org/10.1037/a0014939.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. Science, 328(5977), 456–459.

La Paro, K. M., Hamre, B. K., LoCasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D. M., . . . Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. Early Education and Development, 20, 657–692. https://doi.org/10.1080/10409280802541965.

Lobman, C., McLaughlin, J., & Ryan, S. (2005). Reconstructing teacher education to prepare qualified preschool teachers: Lessons from New Jersey. Early Childhood Research & Practice, 7(2).

McBride, J. W., & Silverman, F. L. (1991). Integrating elementary/middle school science and mathematics. School Science and Mathematics, 91(7), 285–292.

McMurrer, J. (2007). Choices, changes, and challenges: Curriculum and instruction in the NCLB era. Washington, DC: Center on Education Policy Retrieved from: http://www.cep-dc.org/displayDocument.cfm?DocumentID=312

Mercer, N., Dawes, R., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: ways of helping children to use language to learn science. British Educational Research Journal, 30(3), 367–385.

Michaels, S., Shouse, A., & Schweinburger, H. (2008). Ready, set, science!: Putting research to work in k-8 science classrooms. Washington, DC: National Academies Press.

Milner, A. R., Sondergeld, T., Demir, K., Johnson, C., & Czerniak, C. (2012). Elementary teachers' beliefs and needs about teaching science: Examining the impact of pre/post NCLB testing in science. Journal of Elementary Science Education, 23(2), 111–132.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. Educational Researcher, 45(1), 18–35. https://doi.org/10.3102/0013189×16633182

National Association for the Education of Young Children. (2015). NAEYC early childhood program standards and accreditation criteria & guidance for assessment. Retrieved from: http://www.naeyc.org/files/academy/file/AllCriteriaDocument.pdf

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common core state standards. Washington, D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading, its implications for reading instruction (NIH Publication No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: The National Academies Press.

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

Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher- led instruction in preschool settings. Early Education and Development, 22, 970–988. https://doi.org/10.1080/10409289.2010.507496

Newton, L. D., Newton, D. P., Blake, A., & Brown, K. (2002). Do primary school science books for children show a concern for explanatory understanding?. Research in Science & Technological Education, 20(2), 227–240.

NGSS Lead States. (2013). Next generation science standards: For states, by states. Washington, DC: National Academies Press Available online at: www.nextgenscience.org/next-generation-science-standards

Ohio Department of Education. (2012, October). Early learning and development standards. Retrieved from http://education.ohio.gov/Topics/Early-Learning/Early-Learning-Content-Standards

Palincsar, A. S., & Magnusson, S. J. (2001). The interplay of firsthand and text-based investigations to model and support the development of scientific knowledge and reasoning. In S. Carver, & D. Klahr (Eds.), Cognition and instruction: Twenty-five years of progress. (pp. 151–194).Mahwah, NJ: Lawrence Erlbaum.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. Science, 328, 459–463.

Piasta, S. B., Yeager Pelatti, C., & Miller, H. L. (2014). Mathematics and science learning opportunities in preschool classrooms. Early Education and Development, 25, 445–468. https://doi.org/10.1080/10409289.2013.817753

Reinhart, M., Bloomquist, D., Gilbert, A., Strickler-Eppard, L., Czerniak, C., Kaderavek, J., & Molitor, S. (2016). Taking science home: Connecting schools and families through early childhood science activity packs. School Science and Mathematics, 116, 3–16.

Romance, N. R., & Vitale, M. R. (1992). A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade four. Journal of Research in Science Teaching, 29(6), 545–554.

Romance, N. R., & Vitale, M. R. (2001). Evolution of a model for teaching in-depth science in elementary schools: Longitudinal findings and research implications. International Journal of Science Education, 23, 373–404.

Romance, N. R., & Vitale, M. R. (2011). A research-based instructional model for integrating meaningful learning in elementary science and reading comprehension: Implications for policy and practice. In N. L. Stein, & S. W. Raudenbush (Eds.), Developmental cognitive science goes to school. (pp. 127–142). New York, NY: Routledge.

Romance, N. R., & Vitale, M. R. (2012). Expanding the role of K-5 science instruction in educational reform: Implications of an interdisciplinary model for integrating science and reading. School Science and Mathematics, 112(8), 506–515.

Samuels, C. (April 21, 2016). Early-childhood STEM programs take root at White House event. Edweek. Retrieved online at http://blogs.edweek.org/edweek/early_years/2016/04/early-childhood_stem_programs_take_root_at_white_house_event.html

Shymansky, J. A., Yore, L. D., & Hand, B. M. (2000). Empowering families in hands-on science programs. School Science and Mathematics, 100, 48–58.

Snow, C. E. (2002). Reading for understanding: Toward a research and development program in reading comprehension. Santa Monica, CA: RAND.

Sondergeld, T. A., Milner, A. R., Coleman, L. J., & Southern, T. (2011). Lessons from the field: Examining the challenges and successes of a mathematics and science program using acceleration and enrichment for gifted urban middle school students. In C. C. Johnson (Ed.), Secondary STEM educational reform. (pp. 193–195).New York, NY: Palgrave MacMillan.

Tate, W. F., Jones, B. D., Thorne-Wallington, E., & Hogrebe, M. C. (2012). Science and the city: Thinking geospatially about opportunity to learn. Urban Education, 47, 399–433. https://doi.org/10.1177/0042085911429974

Tu, T. (2006). Preschool science environment: What is available in a preschool classroom?. Early Childhood Education Journal, 33, 245–251. https://doi.org/10.1007/s10643-005-0049-8

Tuttle, N., Kaderavek, J. N., Molitor, S., Czerniak, C. M., Johnson-Whitt, E., Bloomquist, D., . . . Wilson, G. (2016). Investigating the impact of NGSS-aligned professional development on PreK-3 teachers' science content knowledge and pedagogy. Journal of Science Teacher Education, 27(7), 717–745. https://doi.org/10.1007/s10972-016-9484-1

What Works Clearinghouse. (2013). What Works Clearinghouse: Procedures and standards handbook (Version 3.0). Washington, DC: US Department of Education. Retrieved from http://whatworks.ed.gov

van Kleeck, A., Vander Woude, J. V., & Hammett, L. (2006). Fostering literal and inferential language skills in Head Start preschoolers with language impairment using scripted book-sharing discussions. America Journal of Speech-Language Pathology, 15, 85–95.

Varelas, M., & Pappas, C. C. (2006). Intertextuality in read-alouds of integrated science-literacy units in primary classrooms: Opportunities for the development of thought and language. Cognition and Instruction, 24, 211–259.

Varelas, M., Pappas, C. C., & Rife, A. (2006). Exploring the role of intertextuality in concept construction: Urban second graders make sense of evaporation, boiling, and condensation. Journal of Research in Science Teaching, 43(7), 637–666.

Varelas, M., Pieper, L., Arsenault, A., Pappas, C. C., & Keblawe-Shamah, N. (2014). How science texts and hands-on explorations facilitate meaning making: Learning from Latina/o third graders. Journal of Research in Science Teaching, 51(10), 1246–1274.

Weiss, H. B., Coffman, J., Post, M., Bouffard, S., & Little, P. (2005). Beyond the classroom: Complementary learning to improve achievement outcomes. The Evaluation Exchange, 11(1), 2–17.

Wenger, E., & Lave, J. (1991). Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives). Cambridge, UK: Cambridge University Press.

Wigfield, A., & Guthrie, J. T. (2000). Engagement and motivation in reading. Handbook of reading research, 3, 403–422.

Wilson, S. M., Lubienski, S. T., & Mattson, S. L. (1996). Where's the mathematics? The competing commitments of professional development. New York, NY: American Educational Research Association.

Yoon, K. S., Duncan, T., Wen-Yu Lee, S., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, Rel 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. Reading Research Quarterly, 39, 347–352.

Zucker, T. A., Justice, L. M., Piasta, S. B., & Kaderavek, J. N. (2010). Preschool teachers' literal and inferential questions and children's responses during whole-class shared reading. Early Childhood Research Quarterly, 25(1), 65–83.

Supporting Information

Additional Supporting Information may be found in the online version of this article.