UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ENHANCING SCIENCE TEACHING THROUGH PROFESSIONAL DEVELOPMENT: IMPACTS OF THE EDSTEMA PROGRAM ON TEACHER KNOWLEDGE, SELF-EFFICACY, AND STUDENT ACHIEVEMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By BRANDI WILLIAMS

Norman, Oklahoma

2025

ENHANCING SCIENCE TEACHING THROUGH PROFESSIONAL DEVELOPMENT: IMPACTS OF THE EDSTEMA PROGRAM ON TEACHER KNOWLEDGE, SELF-EFFICACY, AND STUDENT ACHIEVEMENT

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF EDUCATIONAL LEADERSHIP AND POLICY STUDIES

BY THE COMMITTEE CONSISTING OF

Dr. J. Anthony Luévanos, Chair

Dr. Curt Adams

Dr. Jeffrey Maiden

Dr. Jacob Pleasants

Table of Contents

Abstract

This study examines the impact of Every Day STEM for All (EDSTEMA)—a multi-district professional development (PD) program—on elementary teachers' science content knowledge (SCK), science teaching efficacy (STE), and their students' science achievement (SAA). Grounded in Desimone's (2009) framework for effective professional development, EDSTEMA was designed to incorporate all five critical features: sustained duration across a year of learning, a strong content focus on relevant science topics, coherence with state standards and district goals, active learning through inquiry-based activities and lesson design, and collective participation within school-based cohorts.

Teacher outcomes were measured using selected items from the Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) to assess SCK and the Science Teaching Efficacy Belief Instrument (STEBI-A) to assess STE. Student outcomes were evaluated using Oklahoma Core Curriculum Test (OCCT) scores.

Results showed significant improvement in teacher SCK and STE among EDSTEMA participants. Independent-samples *t*-tests revealed strong, highly powered effects for SCK, while analysis of covariance (ANCOVA) results for STE indicated a small but practical effect, reflecting gains in teaching confidence and practice. Student analyses further revealed that classes taught by EDSTEMA participants scored significantly higher on the OCCT following teacher participation.

Findings suggest that thoughtfully designed PD can enhance teacher knowledge, strengthen efficacy, and improve student achievement. This study extends the literature by showing how growth in SCK and STE is associated with SAA in elementary science, underscoring the importance of investing in sustained, high-quality PD models like EDSTEMA.

List of Abbreviations

Abbreviation	Meaning
ANCOVA	Analysis of covariance
APA	American Psychological Association
AYP	Adequate yearly progress
CG	Comparison group
CoP	Community of practice
CK	Content knowledge
DV	Dependent variable
EDSTEMA	Every Day STEM for All professional development program
ELA	English language arts
FRL	Free or reduced lunch
IRB	Institutional Review Board
IV	Independent variable
MAR	Missing at random
MOSART	Misconception-Oriented Standards-Based Assessment Resources for Teachers
MSP	Math and Science Partnership Grant
NA	Needs assessment
NCLB	No Child Left Behind
NOS	Nature of science
NRC	National Research Council
NSES	National Science Education Standards
NSF	National Science Foundation
NSSME+	National Survey of Science and Mathematics Education Plus
OAS	Oklahoma Academic Skills
OCCT	Oklahoma Core Curriculum Tests

PASS	Priority Academic Student Skills
PCK	Pedagogical content knowledge
PD	Professional development
PL	Professional learning
PLC	Professional learning community
PLN	Professional learning network
PSTE	Personal science teaching efficacy beliefs
RQ	Research question
SAA	Student academic achievement
SCK	Science content knowledge
SI	Summer institute
SoLD	Science of learning and development
STE	Science teaching self-efficacy
STEBI-A	Science Teaching Efficacy Beliefs Instrument for in-service teachers
STEM	Science/technology/engineering/mathematics
STOE	Science teaching outcome expectancy
TG	Treatment group
TSE	Teaching self-efficacy
Year 0	School year prior to participation in EDSTEMA
Year 1	School year of EDSTEMA participation

Chapter 1: Introduction

Science education in the United States has faced persistent and systemic challenges that have threatened the development of scientific literacy among both students and the general public. As scientific understanding has become increasingly vital for informed decision-making and public policy, American students have continued to lag behind their international peers in science achievement. Challenges have included reduced instructional time for science in elementary schools and inequitable access to high-quality instruction. Although professional development (PD) has been widely recognized as essential for strengthening science education, research on the effectiveness of PD programs has been limited. This chapter examines these interrelated issues and highlights how effective professional learning (PL) can enhance both teacher performance and student outcomes.

Importance of Scientific Literacy

America's schools have continued to struggle to produce citizens who are sufficiently scientifically literate. National surveys have revealed that large segments of the population lack foundational science knowledge. The Science and Engineering Indicators Report (Funk & Goo, 2015) has found that more than 40% of American adults were unable to answer basic science questions correctly, underscoring a widespread deficit in scientific understanding.

This lack of literacy is especially concerning because scientific and technological knowledge now permeates nearly every aspect of daily life. The National Research Council (NRC, 2014) has emphasized that understanding science and engineering is essential for informed citizenship:

> A compelling case can be made that understanding science and engineering, now more than ever, is essential for every American citizen. Science, engineering, and the

> technologies they influence permeate every aspect of modern life. Indeed, some knowledge of science and engineering is required to engage with the major public policy issues of today as well as to make informed everyday decisions. (p. 7)

Scientific literacy enables individuals to make thoughtful personal choices and engage with science-related public issues. Beyond its civic value, scientific understanding enriches daily life by fostering curiosity and lifelong growth. International comparisons have reinforced the urgency of this challenge. Results from the Programme for International Student Assessment (PISA) have consistently shown that U.S. students perform below peers in countries such as Finland, Singapore, and South Korea (Darling-Hammond, 2010). These disparities reflect broader systemic issues in how science is taught, valued, and supported across American schools. International comparisons have further underscored the need for improvement.

Challenges in U.S. Science Education

One major reason for Americans' limited scientific understanding has been the reduced time devoted to science instruction in U.S. schools (Karweit, 1983; Funk & Rainie, 2015). Over the past three decades, elementary schools have consistently devoted far less instructional time to science than to mathematics or English language arts (ELA), and this imbalance continues to worsen (Blank, 2013; Trygstad et al., 2013). The decline is especially troubling because elementary school is a pivotal period for developing scientific thinking and curiosity (NRC, 2012).

Early exposure to science builds a foundation for lifelong engagement and problem-solving. Hands-on inquiry builds critical-thinking skills that transfer across disciplines. Without these early experiences, students are less likely to pursue science, technology, engineering, and mathematics (STEM) fields later in life. Indeed, nearly one-quarter of college STEM majors have reported deciding on their field by middle school or earlier (Harris Interactive, 2011). Yet interest in science often declines as early as fourth or fifth grade (Osborne, 2003).

When young learners lack access to quality science instruction, they miss opportunities to develop curiosity, persistence, and reasoning skills that are increasingly essential for navigating a complex, technology-driven world.

Decline in Science Instructional Time

Multiple factors have contributed to the steady decline in elementary science instruction (Council of State Science Supervisors, 2006; Lee & Houseal, 2003), including:

- accountability pressures from federal and state systems;
- larger, more diverse classes;
- shrinking budgets for materials; and
- limited teacher preparation or confidence in teaching science

These factors have made it difficult for schools to maintain consistent, high-quality science instruction.

The problem has deep historical roots. Beginning with reform efforts in the 1980s and 1990s, policymakers prioritized literacy and mathematics as primary indicators of student academic achievement (SAA). This emphasis aimed to improve student performance but inadvertently marginalized other subjects, particularly science. As accountability systems grew more rigid, instructional time and resources were increasingly redirected toward tested areas.

The Rise of Standardized Testing and Accountability Policies

The emphasis on accountability intensified after the passage of the No Child Left Behind (NCLB) Act of 2001, which aimed to raise achievement through rigorous standards-based testing and accountability policies (Hanushek & Raymond, 2005). NCLB required states to create detailed standards, develop proficiency definitions, and evaluate schools based on whether students met *adequate yearly progress* (AYP) targets—both overall and by subgroup—toward the ambitious goal of 100% proficiency by 2014 (Figlio & Loeb, 2011).

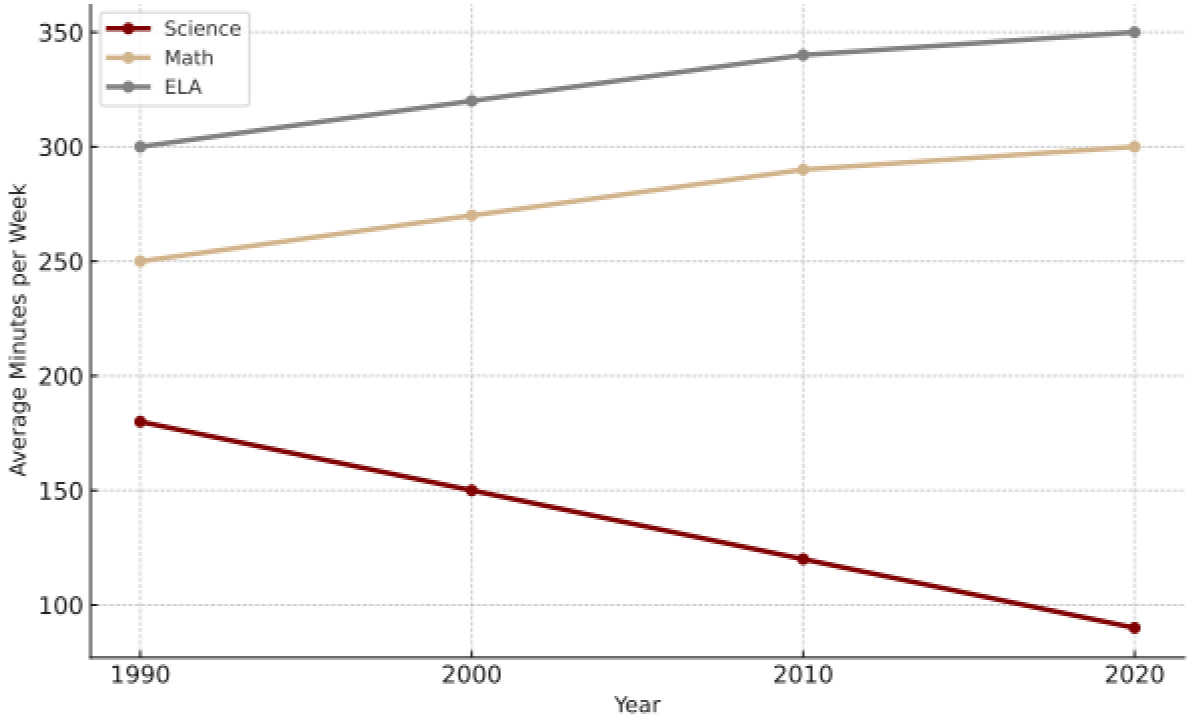
Although well-intentioned, the focus on reading and mathematics left little instructional time for science. During this era, it was estimated that schools had only 9,000 hours available to teach more than 15,500 hours' worth of required material (Marzano, 2003). To meet accountability demands, administrators were forced to make difficult choices about how to allocate limited instructional time. Because science did not count toward AYP, many districts reduced or eliminated science instruction to focus on subjects tied to federal accountability measures. As a result, elementary students often received little to no formal science instruction until middle school. This narrowing of the curriculum marked a significant step backward for equitable and inquiry-based learning.

Consequences of the Shift in Instructional Priorities

These shifts have had lasting consequences. Figure 1 illustrates how, over three decades, time devoted to science steadily declined while time to devoted to English language arts (ELA) and math increased. While the establishment of state science standards under No Child Left Behind (NCLB) represented progress toward consistency, the overwhelming focus on testing accountability had unintended consequences. Teachers—under pressure to meet proficiency targets—often lacked time and resources to engage students in meaningful inquiry. This narrowed the curriculum and encouraged rote coverage of fragmented facts rather than deep, conceptual understanding of scientific phenomena.

Figure 1

Changes in Average Elementary Instruction Time (1990–2020)



Note: This figure illustrates general trends in average weekly instructional time for elementary science, mathematics, and English language arts (ELA) from 1990 to 2020. The steady decline in science instructional time compared with increases in ELA and math reflects shifting educational priorities and accountability pressures that have emphasized tested subjects over science instruction (Banilower et al., 2018).

Overcoming the educational inequities and curricular imbalances that emerged during this period will require sustained effort, targeted investment, and a renewed commitment to science as a core component of elementary education. Districts and states will need to protect dedicated instructional time for science, align accountability systems so that science is not displaced by tested subjects, and ensure equitable access to materials, laboratories, and enrichment opportunities. Equally important, teachers will need sustained, content-focused professional development (PD) that will build science content knowledge (SCK) and teaching self-efficacy (TSE), and that will support the use of inquiry-based, culturally-responsive practices in diverse classrooms. Together, these steps can help restore coherent, concept-rich science learning for all students.

Addressing these systemic challenges will require not only restoring time for science in the curriculum but also confronting the inequities that determine which students have access to meaningful, high-quality science learning and which do not. The following section examines how disparities in resources, teacher preparation, and learning opportunities have deepened the divide in science education across U.S. schools.

Inequities in Science Education

Inequitable access to high-quality science education has continued to widen achievement gaps. In many schools, science instruction has been inconsistent with uneven exposure to core scientific concepts. Those in under-resourced schools have often faced limited instructional time, outdated materials, and teachers with minimal science preparation. These disparities in formal education have prevented many students from developing foundational scientific understanding and critical thinking skills early in their schooling.

Disparities in both formal instruction and informal experiences—such as museum visits or after-school programs—have compounded existing achievement gaps. Children from more affluent families have often benefitted from experiences that extend learning beyond the classroom, while others have had few or no such opportunities. As a result, differences in access to both formal and informal science learning environments have compounded existing disparities in academic achievement (Barton & Coley, 2008; Braun et al., 2009; Darling-Hammond, 2010; National Center on Time and Learning, 2010; Traphagen, 2011).

Impact on Marginalized Student Populations

The consequences of these inequities have been most pronounced for students from historically marginalized populations. Students from low-income backgrounds, students of color, English language learners (ELLs), and students with disabilities have been disproportionately affected by the lack of high-quality science instruction. Schools serving these populations have often faced limited funding, larger class sizes, and fewer qualified teachers, making it difficult to sustain effective science programs. Findings from the 2018 *National Survey of Science and Mathematics Education Plus* (NSSME+) confirmed these disparities: elementary teachers nationwide devoted minimal time to science instruction, and schools serving higher proportions of low-income students were especially likely to lack well-prepared educators and adequate instructional resources (Banilower et al., 2018).

Broader Implications for Equity

These inequities have not only restricted individual academic and career pathways but have also weakened the nation's capacity to cultivate a diverse, skilled STEM workforce capable of addressing complex scientific and societal challenges. Ensuring equitable, inquiry-based science education is not just a matter of fairness—it is a national imperative. Addressing these systemic inequities will require more than additional resources—it will demand sustained investment in teacher learning. The following section examines how high-quality professional development can strengthen teachers' knowledge, confidence, and instructional practices, ultimately expanding access to meaningful science education for all students.

Professional Development for Science Teachers

High-quality professional development (PD) is essential to improving science instruction, yet many teachers have faced significant barriers to accessing it. Professional development (PD) is intended to strengthen teachers' content knowledge (CK) and pedagogical skills, but the quality and effectiveness of such programs have varied widely. Too often, PD experiences have been brief, disconnected from classroom realities, or insufficiently aligned with research-based best practices. Teachers need ongoing, context-specific learning supported over time; without such continuity, even well-designed PD initiatives have rarely produced lasting change in classroom practice.

Federal Policies Which Support Professional Development

To address the need for stronger teacher learning, federal and state governments have long invested in science education reform. From 1985 to 1996, the *Eisenhower Mathematics and Science Education Program* funded professional development for K–12 teachers, followed by the *Math and Science Partnership* (MSP) initiative, which provided competitive grants from 2002 to 2016.

These programs aimed to enhance teacher content knowledge (CK) and instructional quality to raise student academic achievement (SAA) in mathematics and science (United States Department of Education [USDE], 1997; USDE, 2012). Despite these efforts and a large body of research on PD in general, relatively few studies have examined the specific impact of science-focused PD, particularly at the elementary level (NRC, 2013). This lack of evidence has limited understanding of how PD can most effectively improve both teaching and learning in science.

Research Findings

Recent large-scale data confirm these persistent challenges. The 2018 National Survey of Science and Mathematics Education Plus (NSSME+) survey confirmed these challenges. Most elementary teachers have been devoting minimal time to science, and access to high-quality professional development (PD) has been uneven. Although PD has been widely recognized as a pathway for improvement, opportunities have been fragmented, short-term, unevenly distributed across contexts, and disconnected from classroom realities. These findings highlight an urgent need for PD models that are sustained, coherent, and content-focused—features that have been consistently linked to improved teacher knowledge and student learning.

The field of science teacher PD has been strengthened by research that demonstrated how specific features of PD influenced teacher and student outcomes. Such studies provided clear guidance to program developers by identifying which elements contributed most to effectiveness. Teachers, in turn, benefitted from understanding what kinds of professional learning (PL) opportunities were most likely to improve their knowledge, confidence, and instructional practices. Funders and policymakers were able to use this evidence to recognize and invest in promising PD models, ensuring that resources were directed toward initiatives with the greatest potential impact.

Need for This Study

Given these ongoing challenges and evidence gaps, additional research on science teacher professional development (PD) remains critical. This study responded directly to that need by evaluating the long-term, cohort-based PD initiative *Every Day STEM for All* (EDSTEMA), which was designed to strengthen elementary teachers' science content knowledge (SCK) and teaching efficacy (TSE) while improving student outcomes. By examining what makes PD programs effective, this research contributed insights that can inform educators, administrators, and policymakers. Ultimately, these findings aim to support the development of targeted, sustainable PD initiatives that promote teacher growth, advance student achievement (SAA), and foster a more scientifically literate society.

Building on these national efforts and research findings, this study sought to address the persistent gap in understanding how sustained, content-focused PD influenced both teachers and their students. The following section outlines the specific purpose of this investigation and describes how EDSTEMA was designed to meet these needs through a long-term, cohort-based model of PL.

Purpose of the Study

The purpose of this study was to contribute to the growing body of literature on effective science teacher professional development (PD) by evaluating the impact of Every Day STEM for All (EDSTEMA), a PD program which was funded through the Math and Science Partnership (MSP) initiative. EDSTEMA was intentionally designed to enhance classroom science instruction by providing teachers with sustained, research-based learning experiences. This study examined how participation in EDSTEMA influenced both teacher and student outcomes, with particular attention to three key constructs: *science content knowledge* (SCK), *science teaching self-efficacy* (STE), and *student academic achievement* (SAA) in science.

More specifically, this research explored the effectiveness of a sustained, coherent, cohort-based PD model in which teachers engaged deeply with both science content and instructional strategies. Given the increasing recognition that effective PD must be immersive, long-term, and collaborative, this study investigated how these features contributed to teacher growth and, in turn, improvements in student learning. The study also sought to provide evidence of how sustained professional learning (PL) opportunities influenced teachers' confidence and instructional quality, offering insight into the mechanisms by which PD could enhance teaching efficacy and student performance.

The need for programs such as EDSTEMA emerged from the interconnected challenges confronting science education nationwide. As illustrated in Figure 2, reductions in elementary science instructional time have limited students' early exposure to fundamental scientific concepts, contributing to lower scientific literacy, and widening achievement gaps. These systemic issues underscored the urgency of developing and implementing high-quality PD programs that aimed to strengthen teacher knowledge, promote instructional confidence, and foster equitable access to rigorous science learning for all students.

Figure 2

Conceptual Map of Science Education Challenges

Reduced Science Instructional Time

↓

Limited Early Exposure to Science Concepts

↓

Lower Science Literacy and Achievement Gaps

↓

Need for Effective Teacher Professional Development (PD)

↓

EDSTEMA Program Response

Note: This conceptual map illustrates the interrelated challenges contributing to the decline of science education in the United States. Reduced instructional time in elementary science limits early exposure to core concepts, leading to lower scientific literacy and widening achievement gaps. These challenges highlight the need for high-quality, sustained professional development for teachers. The EDSTEMA program represents a targeted, research-based response designed to address these issues through cohort-based professional learning.

Problem Statement

Although professional development (PD) has been widely recognized as essential for improving teaching and learning, there has been limited understanding of how specific PD models influenced both teacher and student outcomes over time. Many PD initiatives have aimed to enhance teachers' content knowledge (CK) and instructional practices, yet relatively few studies have examined their long-term effects, particularly in the context of science education. This gap in evidence has made it difficult for educators, researchers, and policymakers to identify which PD structures most effectively strengthen teaching efficacy and improve student achievement.

Gaps in Existing Research

National data have underscored the urgency of addressing the gaps in science education. The 2018 National Survey of Science and Mathematics Education Plus (NSSME+) revealed that elementary teachers devoted very limited instructional time to science, leaving many students with inconsistent exposure to foundational concepts (Banilower et al., 2018). The same survey found that opportunities for PD have often been fragmented, short-term, and unevenly distributed across schools and districts. These findings illustrated the insufficiency of current PD practices and highlighted the pressing need for sustained, coherent, and content-focused PD models that can more effectively support teacher growth and, ultimately, student learning.

Rationale for the Study

This study addressed the need for research examining long-term, cohort-based, and content-rich professional development (PD) models that support teacher learning and improve student outcomes in science. Every Day STEM for All (EDSTEMA) was developed to fill this gap by providing sustained PL designed to enhance teachers’ science content knowledge (SCK) and science teaching self-efficacy (STE) while promoting student success in science.

By examining the effects of EDSTEMA, this research aimed to deepen the understanding of how comprehensive PD initiatives contributed to teacher development and student achievement. The research offered insights into the PD features most strongly associated with effectiveness in authentic educational contexts—particularly those emphasizing sustained engagement, collaboration, and content focus.

A significant challenge in science education has been understanding how PD can not only improve teachers’ CK and STE, but also lead to meaningful improvements in student academic achievement (SAA). While prior studies have established strong links between teacher effectiveness and student performance, limited evidence existed on how long-term PD programs, such as EDSTEMA, fostered these improvements over time and translated them into enhanced student learning outcomes.

Focus of the Investigation

This study addressed that gap by evaluating the impact of the Every Day STEM for All (EDSTEMA) program on both teacher and student outcomes. Specifically, it investigated how a sustained, coherent, cohort-based professional development (PD) model influenced teachers’ science content knowledge (SCK) and science teaching self-efficacy (STE) and whether these gains corresponded with measurable improvements in student academic achievement (SAA). By analyzing these relationships, the study provided valuable insights into the potential benefits of immersive, long-term, and collaborative PD initiatives. These findings aimed to inform the design of future professional learning (PL) programs that aim to strengthen science education at the elementary level.

Research Questions (RQs)

This study addressed three central research questions regarding participation in the Every Day STEM for All (EDSTEMA) professional development (PD) program:

RQ1: Did participation in EDSTEMA significantly improve elementary teachers' science content knowledge (SCK)?

RQ2: Did participation in EDSTEMA significantly improve elementary teachers' science teaching self-efficacy (STE)?

RQ3: Did participation in EDSTEMA significantly improve student academic achievement (SAA) in science compared to the previous year?

While teacher effectiveness has been shown to play a significant role in student performance, there has been limited research on how specific PD models, such as EDSTEMA, directly influenced SAA over time. By focusing on a sustained, coherent, cohort-based PD program where participants were deeply and actively engaged in both science content and pedagogical strategies, and which encouraged collaboration and continuous reflection, this study aimed to shed light on whether and how such PD programs led to measurable improvements in student learning in science.

Significance of the Study

This research was timely and relevant. It responded to ongoing concerns about the quality and accessibility of science instruction in U.S. schools and added to what was known about effective professional development (PD).

Although PD has generally been shown to improve teacher content knowledge (CK) and science teaching self-efficacy (STE), relatively few studies have examined these features in the context of elementary science or investigated their connection to student academic achievement

(SAA). Because elementary teachers have often been underprepared to teach science, and because PD research has disproportionately emphasized secondary education, a gap remained in understanding how PD functioned in elementary contexts. This study responded directly to that gap by investigating the effects of a sustained, cohort-based PD model designed specifically for elementary teachers.

Contributions to Theory and Practice

Using Every Day STEM for All (EDSTEMA) as a case, this research explored how sustained, coherent professional development (PD) influenced teacher science content knowledge (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) in science. The goal was to clarify how professional learning (PL) influenced classroom practice and, ultimately, student outcomes, thereby contributing to theoretical models of effective PD. Rather than serving solely as an evaluation of a single program, the study examined the relationships among these variables to develop a more comprehensive understanding of the pathways through which PD impacted student learning.

The study applied Desimone's (2009) framework for effective PD—which emphasized duration, content focus, coherence, active learning, and collective participation—within the under-researched context of elementary science and a cohort-based, district-level partnership. This extended the application of a widely cited framework into a context that has received less empirical attention, while generating practical insights for policymakers, PD designers, and school leaders. Ultimately, this research highlighted not only whether PD is effective, but also the factors that contributed to its success, offering guidance for creating more effective, equitable, and sustainable PL experiences.

Implications for Policy and Practice

Evaluating the effectiveness of Every Day STEM for All (EDSTEMA) offered broader lessons about how professional development (PD) could be designed to meaningfully improve both teaching and learning in elementary science. Specifically, it examined how PD features such as sustained duration, content focus, coherence with district goals, and collective participation shaped teacher science content knowledge (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). In doing so, the study provided insights that extended beyond a single program and could inform decisions at multiple levels in the education system. The following sections outline implications of this study for policymakers, district and school leaders, and PD developers and researchers.

Implications for Policymakers

- **Invest in Long-Term, Coherent Professional Development Models.** Effective PD requires sustained duration, collaboration, and alignment with systemic instructional goals—not short-term or fragmented initiatives.
- **Prioritize Funding for Scalable and Evidence-Based Designs.** Findings could guide state and federal funding strategies toward PD structures most likely to yield measurable improvements in teacher growth and student outcomes.

Implications for District and School Leaders

- **Embed Professional Development within Local Contexts and Teacher Cohorts.** Locally-grounded, cohort-based approaches promote coherence, collaboration, and sustained changes in instructional practice.

- **Ensure Leadership Support and Protected Time for Implementation.** Effective PD depends on administrative structures that provide protected time, resources, and support for job-embedded professional learning (PL).

Implications for Professional Development Designers and Researchers

- **Bridge Teacher and Student Outcomes.** This study linked improvements in teacher SCK and STE with gains in SAA, offering a model for integrating teacher- and learner-focused metrics.
- **Test and Refine Theoretical Frameworks.** Applying Desimone's (2009) framework within a cohort-based, district-level partnership advanced understanding of how PD features interact in authentic contexts.

Overall Significance

Ultimately, this research moved beyond asking just whether PD works. By linking teacher and student outcomes, it offered both practical and theoretical guidance for educators, policymakers, and PD designers. These insights strengthened the foundation for creating PL experiences that improve science education across diverse educational settings.

Theoretical and Philosophical Framework

Guided by Desimone's (2009) framework for effective professional development (PD), this study drew from post-positivist and pragmatic traditions. The post-positivist lens emphasized systematic, evidence-based inquiry while acknowledging the limits of educational contexts. The pragmatic perspective complemented this by focusing on generating knowledge that improves practice. Constructivist ideas also informed the design, recognizing that teachers build knowledge and confidence through collaborative, sustained learning experiences (Creswell, 2014).

Limitations of the Study

Despite its contributions, this study had several limitations related to participant selection, measurement tools, and timing of assessments. These factors may have affected the generalizability of findings but also highlighted directions for future research.

Self-Selection Bias and Attrition

Because participation in the Every Day STEM for All (EDSTEMA) program was voluntary, the study lacked random assignment between the treatment (TG) and comparison groups (CG). This self-selection introduced the potential for bias, as participants who chose to enroll in the program may have differed from those who did not in motivation, prior experience, or instructional approach. Additionally, attrition within the CG reduced the overall sample size, which may have limited the generalizability of the results.

Instrument Limitations

The instruments used to measure teachers' science content knowledge (SCK) and science teaching self-efficacy (STE) each carried inherent limitations. Although these tools have demonstrated validity in various educational contexts, they may not have fully captured the complexity of teacher knowledge or efficacy within a cohort-based, sustained PD model such as Every Day STEM for All (EDSTEMA). Similarly, the measure of student academic achievement (SAA) was affected by ongoing revisions to state science assessments during the study period, potentially influencing score comparability.

Scope of Interpretation

Additionally, the participating districts in this study were all suburban systems located near a metropolitan area. As a result, the implementation and outcomes of Every Day STEM for All (EDSTEMA) may differ in rural and urban districts. Rural systems may face challenges such as geographic isolation, small teaching staffs, and limited access to substitutes or release time for extended professional development (PD), whereas large urban districts may contend with higher teacher turnover, competing reform initiatives, and more complex organizational structures. These contextual differences should be considered when applying the findings or adapting the EDSTEMA model to other settings.

These limitations should be considered when interpreting the results and generalizing them beyond the study sample. While they may constrain the external validity of the findings, they do not diminish the study's contribution to understanding how long-term, content-focused PD influences teacher and student outcomes. Instead, they highlight areas for continued inquiry, such as validating adult-focused science knowledge instruments, examining PD effects across diverse settings, and exploring strategies to improve study retention and assessment consistency.

EDSTEMA Overview

The Every Day STEM for All (EDSTEMA) program was created through a partnership among five suburban Oklahoma school districts, two universities, and a research institution with the aim of strengthen elementary science education. District leaders initiated the project after identifying low student performance and limited instructional time in science as persistent issues.

The following sections are built on this foundation, first presenting the rationale for EDSTEMA, including the needs assessment that demonstrated the urgency for a sustained, content-rich professional development model. This is followed by a discussion of the creation and design of EDSTEMA, outlining how the program was structured to address identified gaps in teacher knowledge, efficacy, and student outcomes.

Purpose of the Partnership

The Every Day STEM for All (EDSTEMA) collaboration was designed to provide a systemic response to these challenges by developing a sustained, content-rich professional development (PD) model for elementary teachers. The program sought to strengthen teachers' science content knowledge (SCK) and science teaching self-efficacy (STE), thereby improving instructional quality and ultimately enhancing student learning outcomes in science.

This coordinated effort emphasized shared responsibility among school districts, universities, and research partners, ensuring that PD was grounded in both educational practice and scientific expertise. By aligning local needs with national priorities in STEM education, the partnership positioned EDSTEMA as a model for collaborative, research-based teacher development.

Rationale for EDSTEMA

A voluntary survey of teachers revealed that, on average, the elementary teachers in the partner districts devoted the equivalent of about two 55-min periods per week—or just 21 min per day—to teaching science. This was significantly less than the 90 min per day mandated for English language arts (ELA) instruction. While some teachers devoted more time to science, others devoted no time at all.

Informal interviews with teachers further illustrated this imbalance. Many described how their instructional days were dominated by ELA and math, with science competing for whatever time remained after lunch, recess, social studies, and *specials* (e.g., PE, art, music). Science instruction was often deprioritized to make room for remediation in tested subjects, reflecting the broader systemic pressures of standardized testing. This pattern illustrated the systemic challenges facing elementary schools, where science has been frequently sidelined in favor of subjects more directly tied to accountability measures.

This pattern of devoting limited time to science instruction was not unique to these five districts; it mirrored national trends showing that elementary schools across the country have gradually reduced time dedicated to science in favor of other subjects—particularly math and ELA. Studies showed that elementary schools across the United States had progressively reduced time allocated to science in favor of ELA and mathematics (Blank, 2013; Trygstad et al., 2013). This shift, driven largely by high-stakes testing policies, has had widespread and long-term consequences for science literacy and student preparedness.

As a result, many students entered middle school with limited foundational knowledge of science concepts. Middle school teachers frequently had to reteach basic content that should have been mastered earlier, leaving less time for conceptual development and inquiry-based learning. By high school, teachers faced the compounded effects of this instructional deficit, often spending valuable instructional time reviewing remedial material instead of engaging students in higher-level scientific analysis and application.

Creation of EDSTEMA

This systemic reduction of science instructional time at the elementary level demonstrates the need for comprehensive, research-based professional development (PD) that empowers teachers to integrate meaningful science instruction into their daily practice. In response to these challenges, the five participating districts formed a partnership to strengthen science education at the elementary level. Funded by a $200,000 Mathematics and Science Partnership (MSP) grant, Every Day STEM for All (EDSTEMA) provided sustained, content-rich PD to build teachers' science content knowledge (SCK), instructional confidence, and classroom practices. The resulting EDSTEMA program was designed to equip teachers with the knowledge and strategies needed to deliver high-quality science instruction, thereby giving elementary students a strong foundation in science to support their future learning and academic success.

Summary

In summary, Every Day STEM for All (EDSTEMA) was a collaborative, Math and Science Partnership (MSP) funded professional development (PD) program created to address persistent challenges in elementary science education. By strengthening teachers' content knowledge and instructional confidence, it sought to improve student achievement and long-term engagement with science. The findings from this study have the potential to inform district, state, and national policies on how targeted, well-structured PD can create sustainable improvements in teaching and learning.

Chapter 2: Literature Review

This chapter reviews research on professional development (PD), emphasizing studies that informed the present investigation. While not exhaustive, it highlights key empirical and theoretical works that informed the study’s objectives, situating it within the broader field of teacher learning. The discussion begins with an overview of teacher characteristics such as content knowledge (CK), pedagogical expertise, and self-efficacy (TSE), as well as their influence on student academic achievement (SAA), establishing the foundation for understanding how effective PD can enhance teacher practice and student outcomes. This discussion is relevant to the study because it situates the investigation of PD within a broader understanding of how teacher attributes shape learning outcomes.

The review then shifts to PD more broadly, examining its theoretical foundations, common models, and evidence-based best practices. Special attention is given to the framework proposed by Desimone (2009), which identifies five core features of effective PD—content focus, active learning, coherence, duration, and collective participation. These features have been widely validated across educational contexts and provide a guiding structure for the design and evaluation of PD initiatives. Understanding these foundational elements is essential for evaluating how specific PD programs, such as those designed to strengthen science teaching, can influence teacher knowledge, instructional practice, and student learning outcomes.

Finally, the chapter narrows in scope to focus on PD for science teachers, emphasizing studies that evaluate the effectiveness of programs aimed at improving teachers’ science teaching self-efficacy (STE), science content knowledge (SCK), instructional practices, and student outcomes. This section synthesizes findings from research and meta-analyses to highlight both the potential and limitations of current PD approaches in science education. This synthesis is relevant to the current study because the Every Day STEM for All (EDSTEMA) program was intentionally designed to address many of these limitations—particularly the lack of sustained, content-rich, and coherent PD opportunities for elementary teachers.

Overall, this literature review provided a conceptual and empirical foundation for the study. It not only summarized key findings from prior research but also identified critical gaps, such as the need for longitudinal data linking teacher PD participation to student outcomes and the need for more PD models emphasizing coherence and active learning. These gaps formed the basis for the present study's rationale and underscored its contribution to the field. Ultimately, this chapter establishes the theoretical and practical significance of the investigation by connecting what is known about effective PD to what remains to be explored in the context of science education reform.

Teacher Characteristics That Contribute to Student Learning

A substantial body of empirical research has identified teacher quality as the most significant school-based factor influencing student academic achievement (SAA) (Bolyard & Moyer-Packenham, 2008; Boyd et al., 2008; Ferguson, 1991). This evidence carries important implications for educational policy and classroom practice, suggesting that efforts to improve teacher effectiveness may have the greatest impact on student learning outcomes. Hanushek (1992) emphasized the importance of teacher quality by arguing that the difference in a student's one-year growth in achievement between having a "good" teacher and a "bad" teacher could amount to a full grade-level difference in SAA. Collectively, these findings underscore the critical role teachers play in shaping students' learning trajectories and the necessity of supporting teacher growth at all grade levels.

Research further indicates that teacher quality exerts a stronger influence on SAA than many other commonly cited factors, including class size, per-pupil spending, and teacher salaries. Even variables traditionally viewed as critical to student success, such as family income and minority status, exhibit weaker correlations with SAA than the quality of the classroom teacher (Darling-Hammond, 2000). These findings reinforce the importance of prioritizing teacher quality as a primary lever for educational improvement and equity.

Empirical Evidence From the Tennessee STAR Experiment

A particularly influential investigation of this issue was the Tennessee Class Size Experiment (Project STAR), which examined factors influencing student performance on the Stanford Achievement Test in early elementary reading and mathematics across 79 schools (Nye et al., 2004). The results demonstrated that the individual teacher to whom a student is assigned exerts a greater effect on achievement than the school the student attends.

This finding affirms that teacher quality outweighs most school-level factors in shaping student success. However, the study also revealed an enduring challenge: while it clearly demonstrated that teachers matter, it was unable to identify the specific teacher characteristics responsible for this impact. This gap underscores the need to investigate the attributes—both cognitive and affective—that contribute to high-quality teaching and how these can be cultivated systematically.

Implications for Professional Development and Teacher Growth

Addressing these questions requires a closer look at the components of teacher effectiveness and the mechanisms by which they develop. In particular, professional development (PD) serves as a key vehicle for enhancing teachers' science content knowledge (SCK) and science teaching self-efficacy (STE)—two interrelated domains that have been consistently linked to improved student outcomes. Understanding how these attributes evolve through targeted PD experiences is therefore essential for designing initiatives that foster meaningful and sustained improvements in teaching practice.

Building on this foundation, the following sections examine three interrelated dimensions that inform teacher quality and its development. First, teacher preparation is considered as a foundation for equipping educators with the knowledge and skills necessary for effective practice. Next, the discussion turns to teacher self-efficacy, a critical factor influencing how teachers approach instruction and persist through challenges. Finally, the role of PD is explored, with particular attention to how it can strengthen teachers' SCK and STE to improve student outcomes.

Teacher Preparation

Several studies have sought to identify specific teacher characteristics that influence student achievement. For instance, Akiba et al. (2007) analyzed data from the *Trends in International Mathematics and Science Study* (TIMSS) across 46 nations and found that countries with stronger measures of teacher quality demonstrated higher mathematics achievement. Students achieved at the highest levels when taught by teachers who were fully certified, had at least three years of experience, and had majored in mathematics or mathematics education. These findings underscore the central role of teacher preparation and credentialing in shaping student success.

Similarly, Fetler (1999) analyzed the records of 1.3 million students across 795 California high schools and found that teacher experience, certification and education level were significant predictors of student performance on the Stanford Achievement Test in mathematics. This large-scale analysis reinforced the idea that teacher qualifications affect student outcomes, but also highlighted a need to explore how teaching practices and classroom strategies translate these qualifications into measurable gains in learning.

Building on this, Clotfelter et al. (2007) analyzed data from North Carolina elementary schools and concluded that teacher experience and certification positively influenced student academic achievement (SAA). They further found that teachers' performance on licensing exams contributed significantly to student gains, suggesting that both content mastery and formal preparation matter for achievement.

More recent research continues to affirm the importance of teacher preparation while broadening the understanding of what constitutes teacher quality. Darling-Hammond (2022) and Zeichner et al. (2024) emphasized that effective preparation programs must integrate the *science of learning and development* (SoLD)—a framework connecting content expertise, knowledge of how students learn, and equity-centered practice. Their work highlights that preparation is most impactful when it blends disciplinary depth with pedagogical skill and sustained clinical experience.

Similarly, Rhodes and Marder (2024) found that teachers from comprehensive, university-based preparation programs produced higher student learning gains than those from alternative routes, reinforcing the long-term value of structured preparation that includes both coursework and mentored teaching. These findings demonstrate that the strength and coherence of teacher-preparation programs remain critical determinants of instructional quality and student success.

Teacher Content Knowledge and Pedagogical Preparation

The importance of teacher content knowledge (CK) for improving student academic achievement (SAA) has been repeatedly demonstrated in the literature. Monk (1994), using data from the National Science Foundation's (NSF) Longitudinal Survey of American Youth, examined the relationship between teachers' subject-matter preparation and student performance. Results showed that the number of subject-area courses completed by teachers was positively associated with SAA on the *National Assessment of Educational Progress* (NAEP). Moreover, coursework in pedagogy was also found to contribute positively to student performance, highlighting the value of both subject matter knowledge and effective teaching strategies in improving SAA.

Recent research further elaborates this connection. Darling-Hammond (2022) and Zeichner et al. (2024) argue that teacher education must move beyond simple content mastery toward developing *adaptive expertise*, or the ability to apply CK flexibly in diverse instructional contexts. Likewise, the National Council on Teacher Quality (NCTQ, 2021) review found that programs offering strong clinical practice and explicit preparation in classroom management produced teachers who demonstrated greater instructional readiness and self-efficacy. These studies suggest that CK and pedagogical content knowledge (PCK) are mutually reinforcing; teachers must understand both what to teach and how to support student understanding effectively.

Complementary evidence also emerges from science-specific studies. While most studies on teacher characteristics and SAA have used longitudinal data from standardized mathematics and English language arts (ELA) achievement tests, smaller-scale investigations have also yielded valuable insights.

For example, Magnusson et al. (1992) investigated the impact of teacher science content knowledge (SCK) on student understanding of heat energy and temperature among six eighth-grade teachers. Through post-instruction interviews with teachers and 22 students per class, the researchers found that teachers with stronger conceptual understanding of heat and temperature produced students with deeper and more accurate scientific conceptions. This direct link between teacher SCK and student understanding emphasizes the importance of disciplinary expertise, particularly in science education, where misconceptions are common and conceptual precision is essential.

Certification and Teaching Effectiveness

Studies have also consistently demonstrated that teacher certification is an important indicator of teaching effectiveness and student academic achievement (SAA). Hawk et al. (1985) analyzed Stanford Achievement Test data and found that students taught by certified math teachers outperformed those taught by non-certified teachers. Similarly, Darling-Hammond et al. (2005) examined student data from Houston and reported that teacher certification status, experience, and degree level were all significant predictors of SAA in both reading and mathematics. The researchers found that certified teachers had the greatest impact on SAA. The authors argued that certification served as a proxy for a constellation of other key variables—including content knowledge (CK), pedagogical content knowledge (PCK), classroom management, instructional design and implementation, and collaboration skills—that collectively contribute to instructional quality.

Contemporary studies have expanded this view. For example, VanLone (2022) examined the impact of pandemic-related disruptions on teacher preparation and found that preservice teachers who experienced reduced or virtual student-teaching opportunities reported lower self-efficacy and readiness to manage classrooms.

This highlights how consistent, authentic field experiences, which are a hallmark of effective certification programs, are essential for teacher confidence and long-term success.

Collectively, these findings reinforce that certification remains meaningful not merely as a credential but as a representation of teachers' accumulated content knowledge, pedagogical training, and professional confidence.

Implications for Professional Development

While this body of evidence demonstrates the clear importance of teacher characteristics, it also underscores the need for further research into how these qualities influence teaching effectiveness. Understanding which characteristics most strongly shape instructional practice remains a key challenge for researchers and practitioners alike.

Professional development (PD) serves as a critical mechanism for bridging this gap. High-quality PD can enhance teachers' content knowledge (CK), pedagogical content knowledge (PCK), and teaching self-efficacy (TSE) while promoting reflective and adaptive instructional practices (Darling-Hammond, 2022; Zeichner, 2024). Programs that align with principles of the science of learning and development (SoLD) by combining deep disciplinary knowledge with active learning, collaboration, and ongoing feedback, are particularly effective in sustaining growth. Investigating how PD strengthens these interconnected domains provides valuable insight into how schools can foster teacher growth and, in turn, improve student outcomes. This connection lays the groundwork for the next section, which explores TSE as another key dimension of teacher effectiveness and a central construct in understanding the mechanisms through which PD influences both teaching and learning.

Self-Efficacy

Teacher self-efficacy (TSE) is a core construct in understanding teacher effectiveness and its influence on student academic achievement (SAA). Bandura (1977) defined self-efficacy as an individual's belief in their capacity to successfully organize and execute the actions required to achieve desired outcomes in a specific domain. He noted that these beliefs strongly influence motivation, resilience, and performance. Within education, TSE refers to teachers' confidence in their ability to promote student learning through effective instruction, classroom management, and student engagement.

Research has consistently shown that teachers with higher TSE employ more effective instructional practices and have students who perform at higher levels than those taught by teachers with lower TSE (Ashton, 1984; Bandura, 1977; Caprara et al., 2006). The influence of TSE is especially critical in challenging subject areas such as science, where teacher confidence can determine not only teachers' instructional decisions but also the depth of inquiry, degree of experimentation, and classroom climate that supports scientific thinking. This is relevant to the present study because understanding how TSE operates within science instruction provides essential context for examining the outcomes of professional development (PD) programs like Every Day STEM for All (EDSTEMA), which explicitly aim to strengthen teachers' science confidence and practice.

Interconnections Between Content Knowledge and Self-Efficacy

The relationship between content knowledge (CK) and teaching self-efficacy (TSE) has been reaffirmed across multiple studies (deLaat & Watters, 1995; Lakshmanan et al., 2011; Sanders & Rivers, 1996; Swanson, 2010; Tinnin, 2000; Wright et al., 1997). Collectively, this body of evidence indicates that increases in CK are often associated with corresponding gains in TSE, which are in turn linked to higher student academic achievement (SAA). More recent work suggests a pathway toward positive change in which growth in CK is related to stronger TSE and, in some contexts, to more effective instruction and improved student outcomes.

This pattern underscores why the present study focuses on professional development (PD) interventions that intentionally target both cognitive (CK) and affective (TSE) dimensions of teacher learning. Content-rich, sustained PD offers teachers structured opportunities to deepen their understanding of key scientific ideas while also engaging in instructional practices that align with those ideas. As teachers work with challenging content, analyze student thinking, and refine their instructional approaches, changes in their knowledge and beliefs are likely to occur together rather than in isolation.

Taken together, these findings highlight an important pattern of relationships among knowledge, confidence, and outcomes: growth in CK is associated with increases in TSE, which are in turn related to stronger instructional effectiveness and SAA. This pattern provides the rationale for examining EDSTEMA as a PD model that integrates both CK enhancement and efficacy-supporting experiences, reflecting evidence that sustained, content-focused PD is linked to improvements in teacher practice and student learning outcomes and may contribute to a broader pathway toward positive change.

Factors Influencing the Development of Self-Efficacy

The development of teaching self-efficacy (TSE) is shaped by a range of factors, including prior teaching experience, mentorship, and ongoing professional development (PD) (Guskey, 2002). Teachers who participate in high-quality PD programs are more likely to build the confidence and expertise needed to address challenging subjects such as science. These findings underscore the critical role of teacher training and support in enhancing TSE and science teaching self-efficacy (STE) and improving student outcomes.

Subsequent research has expanded this view by emphasizing the importance of contextual and systemic influences. Studies show that self-efficacy evolves over time through supportive professional environments that encourage reflection, shared problem-solving, and ongoing learning (Bandura, 1997; Tschannen-Moran & Hoy, 2001).

For example, when teachers engage in structured collaborations such as mentoring relationships, peer observation, or participation in professional learning communities, they report stronger confidence in their instructional decision-making and a greater sense of agency in managing classroom challenges. These experiences contribute to a sustained and adaptive sense of self-efficacy that extends beyond individual classrooms.

Professional development (PD) models that combine content-rich instruction with structured collaboration—such as lesson study, inquiry-based learning teams, or ongoing coaching—appear particularly effective in fostering TSE (Caprara et al., 2006; Holzberger et al., 2020). These approaches allow teachers to experience success through mastery, feedback, and reflection, all of which reinforce self-efficacy development. This is directly relevant to the present study, which investigates the impact of a PD program intentionally structured around these same principles of active, collective, and coherent professional learning designed to enhance both teacher confidence and instructional quality.

In particular, Every Day STEM for All's (EDSTEMA's) emphasis on content-rich inquiry tasks (mastery experiences), observing university faculty and colleagues modeling practice (vicarious experiences), and structured feedback and encouragement from facilitators and peers (social persuasion), within a collegial and low-risk environment, closely aligns with Bandura's (1977) theorized sources of efficacy formation.

Implications for Professional Development and Student Achievement

Teacher self-efficacy (TSE) plays a critical role between teacher knowledge and student outcomes. When teachers believe in their capacity to teach effectively, they are more likely to take instructional risks, persist through challenges, and foster learning environments that encourage student inquiry and problem-solving. The literature consistently demonstrates that strengthening teachers' content knowledge (CK) is an important pathway associated with higher teaching self-efficacy (TSE) and, in many studies, with higher student academic achievement (SAA) (Bandura, 1977; Darling-Hammond et al., 2022; Tschannen-Moran & Hoy, 2001). In this body of work, CK, TSE, instructional practice, and student outcomes are linked as part of a broader network of influences rather than a single proven causal chain. Programs like Every Day STEM for All (EDSTEMA)—which utilize research-based best practices in professional development (PD)—represent evidence-informed models that may contribute to sustainable improvements in both teacher and student outcomes, although further research is needed to clarify the specific mechanisms involved.

Ultimately, a teacher's belief in their ability to teach science effectively is not only a psychological construct but also a driver of instructional quality. Strengthening TSE through research-based PD serves as a bridge between teacher CK and SAA, underscoring why this study's focus on PD outcomes in SCK, STE, and SAA contributes meaningfully to the ongoing discourse on improving science education.

Professional Development

Given the significant impact of teacher characteristics on student academic achievement (SAA), substantial resources have been devoted to professional development (PD) as a mechanism for improving teacher quality and instructional effectiveness. Guskey (2009) estimated that American schools invest approximately $20 billion annually in PD. These expenditures are well-justified, as research consistently demonstrates that well-designed PD enhances teacher content knowledge (CK), strengthens instructional practices, and improves classroom learning outcomes (Buczynski & Hanson, 2010; Carpenter et al., 1989; Hauck, 2012; Supovitz et al., 2000).

Moreover, numerous studies have found a direct positive relationship between effective PD and SAA (Borko, 2004; Buczynski & Hanson, 2010; Carpenter et al., 1989; Desimone et al., 2002; Fields et al., 2012; Heller et al., 2012; Yoon et al., 2007). Collectively, these findings underscore the critical role PD plays in equipping teachers to meet evolving educational goals and in ensuring that investments in teacher learning translate into meaningful gains for students. This is relevant to the present study because the Every Day STEM for All (EDSTEMA) program was explicitly designed around these evidence-based principles of effective PD to strengthen teachers' science content knowledge (SCK) and science teaching self-efficacy (STE), ultimately improving student outcomes.

Persistent Gaps and National Context

National survey data reinforce these conclusions while also highlighting persistent challenges. The 2018 National Survey of Science and Mathematics Education Plus (NSSME+) is the largest nationally representative snapshot of science, technology, engineering, and mathematics (STEM) education in the United States. It examines teacher characteristics, instructional practices, professional development, and issues of equity (Banilower et al., 2018).

Its findings revealed that elementary teachers devote very limited instructional time to science, and inequities in resources and teacher preparedness are particularly pronounced in schools serving economically disadvantaged populations. Moreover, while professional development (PD) is widely offered, the survey reports that opportunities are often fragmented, short-term, and unevenly distributed across contexts. These shortcomings underscore the need for PD models that are sustained, coherent, and content-focused—the very gaps that Every Day STEM for All (EDSTEMA) aimed to address.

This study responded directly to these national findings by examining the impact of a sustained, coherent PD model on teacher science content knowledge (SCK), teaching self-efficacy (TSE), and student academic achievement (SAA), while promoting equitable access to high-quality science instruction.

Empirical Foundation of Effective Professional Development

To understand which factors contribute to the effectiveness of teacher professional development (PD), Garet et al. (1999) conducted one of the first large-scale empirical investigations comparing PD characteristics and their impact on teacher learning. Emerging from an evaluation of the federally funded Eisenhower Mathematics and Science Education Program, the study sought to analyze how specific PD features were associated with improvements in teaching practice.

Garet et al. (1999) distinguished between *structural features* (duration, collective participation, and form/type of activity) and *core features* (content focus, active learning, and coherence). By surveying a random national sample of teachers and analyzing relationships among these characteristics related to instructional changes, the researchers identified the conditions most strongly associated with meaningful teacher learning. This work provided an evidence-based foundation for understanding what makes PD impactful and offered a framework for assessing program quality. This foundational work directly informed this study's PD model, which utilized these structural and core features in its design to maximize teacher learning and transfer to classroom practice.

Desimone's Framework for Effective Professional Learning

Building upon Garet et al.'s (1999) findings, Desimone (2009) synthesized prior research into a widely recognized framework that defines five critical features of effective professional development (PD):

- content focus,
- active learning,
- coherence,
- sustained duration, and
- collective participation.

Each of these features contributes to PD's overall impact on teacher learning, instructional quality, and ultimately, student academic achievement (SAA). Desimone's model advanced the field by providing a cohesive conceptual structure through which researchers and practitioners can evaluate PD initiatives. Rather than viewing PD as a series of disconnected workshops or isolated training events, her framework emphasizes the importance of designing professional learning experiences that are strategically aligned with instructional goals, grounded in subject-matter content, and supported over time through collaboration and reflection.

Desimone's framework has since been validated and expanded through numerous empirical studies. For instance, Yoon et al. (2007) found that PD programs incorporating these five features were significantly more likely to produce measurable improvements in student learning. Similarly, Buczynski and Hanson (2010) demonstrated that sustained, content-rich PD improved teachers' science content knowledge (SCK) and teaching self-efficacy (TSE), leading to higher-quality science instruction. More recently, Darling-Hammond et al. (2017) affirmed that PD is most effective when it integrates these same elements, especially when teachers engage actively in designing, applying, and reflecting on instructional practices within authentic classroom contexts.

Taken together, these studies confirm that the quality—not the quantity—of professional learning (PL) matters most. PD that is coherent, sustained, and embedded in teachers' everyday work fosters deeper conceptual understanding and supports meaningful instructional change. This is directly relevant to the present study because Every Day STEM for All (EDSTEMA) was intentionally designed to embody these five features, thereby aligning with research-based criteria for effective PD.

The following subsections explore each of Desimone's five features in greater detail. This discussion provides the theoretical and empirical foundation for understanding how and why PD influences teacher outcomes, establishing the framework upon which the EDSTEMA program was developed and evaluated.

Content Focus

Content refers to the subject matter of the professional development (PD). A large body of research supports a content-centered design for PD, demonstrating that initiatives aimed at strengthening teachers' subject-matter understanding lead to measurable improvements in both instruction and student achievement (Carpenter et al., 1989; Fennema et al., 1996; Hiebert et al., 1996; Rhine, 1998; Shulman, 1987). When PD increases teachers' conceptual understanding of core content, it enhances their ability to engage students in higher-level reasoning, inquiry, and problem-solving. However, content-focused PD is most effective when paired with pedagogical strategies.

Doyle's (1990) research on teacher knowledge illustrates this intersection, showing that teacher understanding of content, pedagogical strategies, and student thinking interact dynamically in the classroom. Consequently, PD that integrates content knowledge (CK) and pedagogical content knowledge (PCK) is more likely to influence teacher practice and improve student learning outcomes.

Empirical studies reinforce this conclusion. Garet et al. (1999; 2001) found that content-focused PD alone could hinder teacher practice, whereas PD that strengthened both content and pedagogical skills had a strong positive impact. Building on this foundation, Desimone (2009, 2011) synthesized prior research to develop a widely cited framework for effective PD. Her framework emphasizes not only a focus on content but also the importance of active learning, coherence with teachers' existing knowledge and context, sustained duration, and collective participation.

Further research supports these findings across STEM education contexts. Banilower et al. (2007) found that STEM-focused PD, designed to emphasize both content and pedagogy, led to substantial gains in teacher confidence, instructional quality, and student achievement. Similarly, Yoon et al. (2007) and Guskey and Yoon (2009) demonstrated that PD that focused on strengthening both CK and PCK had a direct positive impact on teacher practices. Rowland et al. (2005) advanced this understanding through their Knowledge Quartet model, which posits that effective teaching requires teachers not only to understand the subject matter but also to know how to communicate it in ways that promote student understanding. These frameworks collectively highlight the importance of PD that unites content and pedagogy rather than treating them as separate domains.

While content-focused PD has been consistently shown to be effective, challenges such as teacher resistance or insufficient school resources can hinder successful implementation.

Kennedy (1998) and Darling-Hammond et al. (2017) note that teacher resistance, limited time, and insufficient school resources can hinder the long-term impact of PD efforts.

PD programs must consider these barriers—along with other factors like school culture, leadership, and collaborative environments—to ensure that PD is both relevant and sustainable. For PD to be sustainable, it must be relevant, contextually grounded, and supported by collaborative school cultures. Desimone and Garet (2011) emphasize that effective PD should align with teachers' goals, encourage professional collaboration, and include follow-up structures to ensure continuity and transfer to practice.

Ultimately, focusing on content has proven to be one of the most powerful levers for improving instructional quality. Yet, it is most effective when integrated with pedagogical knowledge and delivered within supportive professional environments. This is relevant to the present study because EDSTEMA was intentionally designed as content-focused PD that also incorporated pedagogical modeling, active engagement, and collaborative reflection—key factors shown to promote sustained improvements in teacher knowledge, self-efficacy, and student achievement.

Active Learning

Active learning in the context of professional development (PD) requires teachers to apply new content and strategies, enhancing their ability to translate learning into effective classroom practices (Penuel et al., 2007). This approach, grounded in *constructivist learning theory*, views learning as an active process whereby individuals construct new understanding through direct interactions with the world and those who inhabit it (Inhelder & Piaget, 1958; Vygotsky, 1934).

In the context of PD, active learning ensures that teachers are not merely passive recipients of knowledge but are actively engaged in applying, reflecting, and refining new teaching methods. Research consistently shows that PD programs emphasizing active engagement lead to stronger instructional practices and improved teacher confidence (Lieberman, 1995; Penuel et al., 2007). By participating in activities that simulate real classroom scenarios, lesson design activities, and peer collaboration, teachers can more effectively bridge the gap between theory and practice, ultimately improving student outcomes.

According to Darling-Hammond (1995), effective PD should engage teachers both as learners and as educators. This dual engagement allows educators to face the challenges and uncertainties that arise when learning new content and strategies. Engaging as both learner and instructor fosters empathy, reflective practice, and deeper pedagogical insight.

Darling-Hammond's perspective aligns closely with Vygotsky's (1978) concept of the *zone of proximal development* (ZPD), which posits that learning is most effective when individuals receive support while facing challenges. By integrating active learning and reflection, PD becomes a space where teachers—as learners—can explore and experiment with new methods, and as teachers, apply these methods in their classrooms while continually refining their practices.

Empirical studies have validated these principles. Garet et al. (1999; 2001) found that PD providing teachers with opportunities for hands-on work and direct classroom application was more likely to influence teacher practices than PD focused primarily on theoretical or lecture-based learning. For example, activities such as lesson planning, collaborative teaching, and case studies that simulate classroom challenges provide teachers with concrete experiences they can use when implementing new techniques. These hands-on experiences enable teachers to better understand and apply the strategies they are learning. These findings underscore that PD must not only introduce new content but also provide structured opportunities for teachers to apply strategies in real-world settings.

Further emphasizing the importance of implementation, Guskey (1986) argued that active learning should not end once the teachers leave the PD session. Rather, teachers should be tasked with implementing what they learn through PD in their classrooms, because teachers' attitudes and beliefs toward instructional innovation evolve only after they see success implementing newly learned strategies from PD.

This process aligns with Bandura's (1977) theory of self-efficacy, which posits that confidence grows when individuals experience positive results from their actions. Therefore, PD programs must create feedback loops that allow teachers to test new ideas, evaluate outcomes, and adjust their instruction accordingly. Through this iterative cycle of learning, applying, and reflecting, teachers internalize new practices and develop confidence in their ability to sustain them.

Active learning is also strengthened when situated within a *community of practice* (CoP) where teachers collaborate, share experiences, and support each other's growth (Wenger, 1999). Collaborative models of PD, such as peer coaching or professional learning communities (PLCs), have proven to be effective in fostering ongoing reflection and application of new strategies. Teachers working together in these settings can engage in substantive discourse about their teaching practices, brainstorm solutions to challenges, and celebrate successes. This collaborative environment reinforces learning while also promoting a sense of collective responsibility for improving teaching and learning outcomes.

Active learning also aligns with experiential learning theory (Kolb, 1984), which asserts that learning occurs when individuals engage in a continuous cycle of experiencing, reflecting, thinking, and acting. Teachers can move through this cycle by actively participating in PD activities, reflecting on their own experiences, conceptualizing the new information, and then applying it in their classrooms. As they reflect on their experiences, they progress through the cycle and continually refine their practices.

Additionally, PD that incorporates active learning has been linked to improved student outcomes. Penuel et al. (2007) found that when teachers participated in PD that allowed them to actively engage with new material or practices, SAA improved as a result. This connection highlights the importance of high-quality PD as a tool for teacher improvement and student learning. Yoon et al. (2007) found that PDs that included sustained, active engagement—rather than one-off sessions—led to greater teacher effectiveness and student performance. Long-term engagement ensures that teachers have the time to hone new strategies and experience success in their use with students.

In summary, active learning is an important factor in the development of effective PD because it provides teachers with opportunities to apply new content and strategies in meaningful ways. As teachers actively engage with the material and practice new approaches, they are better equipped to make lasting improvements in their practice. However, to ensure that the benefits of active learning in PD are sustained, programs must provide teachers with continuous opportunities to reflect, implement, and refine their practice. This ongoing cycle of learning, doing, and thinking fosters deeper professional growth and improves both teacher effectiveness and student outcomes. This is directly relevant to the present study because Every Day STEM for All (EDSTEMA) is intentionally designed to incorporate active learning experiences, ensuring that teachers can translate professional learning into enduring improvements in both practice and student achievement.

Coherence

Coherence refers to how well a professional development (PD) experience relates to teachers' other professional learning (PL), and how well it aligns to standards, assessments, and district or school goals. A key feature of coherence is the integration of PD into the PL goals of teachers, so that new knowledge and skills build upon previous learning rather than being isolated experiences. When PD is coherent, it provides a continuous and connected learning pathway for teachers, fostering deeper learning and stronger instructional practices. Aligning PD with standards and assessments ensures that it remains relevant to teachers' immediate professional needs while also supporting broader educational goals (Desimone, 2011).

Ultimately, coherence ensures that PD content is directly applicable to teachers' daily work, which increases the likelihood of meaningful implementation of new strategies in the classroom. This is relevant to the present study because Every Day STEM for All (EDSTEMA) was designed to align directly with district science standards and curriculum frameworks, enabling teachers to connect new learning to their daily classroom practice and to existing instructional expectations.

Coherence has been shown to positively influence teacher practice. For example, Grant et al. (1996) and Lieberman (1995) found that when PD was aligned with teachers' instructional goals and the curriculum, it was more likely to produce meaningful changes in classroom practice. Similarly, Garet et al. (1999; 2001) reported that teachers whose PD was coherent were significantly more likely to change their practices compared to those whose PD was fragmented or disconnected.

These findings suggest that coherence plays a critical role in ensuring that PD is not only relevant but also effectively translates into sustained improvements in teaching and learning. In the context of the current study, this principle guided the design of EDSTEMA's modules, which were intentionally structured to connect professional learning to ongoing classroom initiatives, pacing guides, and assessment practices.

From these results, researchers concluded that positive changes in teaching and learning occurred when teachers experienced consistent, high-quality PD that was well-aligned with both their personal goals and the broader school and district objectives. This highlights the importance of coherence in creating continuous and focused learning experiences for teachers, fostering an environment where change is not an isolated event but rather sustained over time.

However, effective, coherent PD requires substantial investment in resources, especially time and money. It requires engaging a cohort of teachers for a sustained period in activities with a clear focus on content and pedagogy, as well as alignment with standards and assessments. This integration of content, pedagogy, and context is essential in creating an effective learning experience that is aligned with teachers' practice and goals.

Leithwood et al. (2004) argued that such coherence ensures teachers are not only learning new strategies but also understanding how these strategies align with broader educational objectives. This alignment helps teachers to see the relevance of what they are learning, increasing the likelihood that they will continue to implement it effectively in their classrooms. For the EDSTEMA study, this is significant because its sustained, cohort-based model was intentionally designed to support coherence through repeated contact sessions, structured collaboration, and continuous alignment with science standards.

Despite its importance, coherence can be difficult to achieve. Many PD programs fail to provide consistent or contextually relevant learning experiences. For example, Hargreaves (2014) pointed out that fragmented PD, which is isolated and unaligned with the broader school context, fails to create lasting change. Furthermore, time constraints and competing priorities may make it difficult for teachers to commit to ongoing, coherent PD efforts.

These challenges underscore the need for a carefully structured PD model that includes ongoing support, follow-up sessions, and opportunities for teachers to share their experiences and reflect on their practice. EDSTEMA's design addressed these challenges by embedding follow-up sessions and collaborative reflection opportunities into the PD sequence, ensuring that learning experiences remained connected, supported, and meaningful over time.

Finally, coherent PD does not just benefit teachers. Research by Penuel et al. (2007) and Yoon et al. (2007) showed that when PD was aligned with standards, assessments, and broader school goals, it had a direct positive impact on student academic achievement (SAA). This alignment ensures that teachers' growth is closely connected to improved teaching practices, which in turn benefits students. In this way, the coherence of PD experiences plays a critical role in shaping the broader educational outcomes for students, creating a cycle of continuous improvement in both teaching and learning. In the present study, coherence serves as a key mediating construct linking EDSTEMA's design features to its measured outcomes in teacher science content knowledge (SCK), science teaching self-efficacy (STE), and SAA.

In conclusion, coherence in PD is a key factor in determining whether it leads to sustained improvements in teaching. When PD is well aligned with teachers' needs, standards, assessments, and school goals, educators are more likely to adopt new practices that bring meaningful change to their classrooms. As research shows, investment in coherent PD supports not only teacher professional growth but also improved student learning. Beyond its immediate impact, coherence creates a foundation for other critical PD features—such as active learning and collective participation—by ensuring that teachers engage in experiences that are relevant, connected, and sustainable.

Sustained Duration

The *duration* of PD refers to the length of time over which the program is delivered. Numerous studies have suggested that PD programs with a substantial number of hours and sustained over a period of time were more effective than those with shorter durations (e.g., Hauck, 2012; Hiebert, 1996; Lieberman, 1995; Loucks-Horsley et al., 2003; Penuel et al., 2007; Scott et al., 2014; Weiss et al., 1998).

Garet et al. (1999, 2001) found that longer PD programs positively affected teacher practices, particularly when they extended over weeks or months rather than being condensed into a short time frame. This extended period allowed teachers to absorb new content, apply it in practice, reflect, and refine their teaching strategies.

Additionally, longer PD programs provided more opportunities for ongoing support and collaboration, which in turn can foster sustained changes in teacher practice and greater improvements in student learning (Desimone, 2011). This is directly relevant to the present study because EDSTEMA was designed as a sustained, multi-phase PD experience, enabling teachers to return to their classrooms between sessions, apply new learning, and then reflect collaboratively in subsequent meetings.

Yoon et al. (2007) conducted a meta-analysis of over 1,300 studies on teacher PD and identified only nine that met the standards outlined by the *What Works Clearinghouse* (Institute of Education Sciences, 2008), an organization which highlights effective educational programs.

The researchers found that programs offering less than 14 hr of PD had no statistically significant impact on SAA. In contrast, programs with more than 14 hr of PD demonstrated significant positive effects on SAA, especially when paired with follow-up activities that reinforced teacher practice.

These findings suggest that PD which is both comprehensive in scope and sustained in duration enables teachers to fully integrate new knowledge into their practice and improve student outcomes. In alignment with these findings, EDSTEMA's extended duration and built-in follow-up components were intentionally structured to promote regular application and reflection, supporting the transfer of new knowledge into classroom practice.

A re-analysis by Guskey and Yoon (2009) further reinforced these findings, confirming that the most effective PD programs required more than 30 clock hr. Desimone (2011) recommends at least 20 contact hr for PD to be effective, while Corcoran et al. (2003) found 50 to 80 hr of PD was most impactful, providing sufficient time for teachers to build mastery over new concepts. Smith and Neale (1989) suggested that PD ideally should extend over a period of up to three years, allowing teachers to develop deep, sustained changes in their instructional practices. Extended PD programs often include periodic check-ins, ongoing learning communities, and reflection opportunities that help solidify the skills and knowledge acquired during the sessions.

However, Kennedy (1998) cautioned that simply increasing the duration of PD is not always beneficial. If the content or delivery of the program is ineffective, increasing the number of hours may lead to teacher disengagement and burnout, rather than improvements in teaching quality. Thus, the effectiveness of PD programs depends not only on duration but also on the quality of content, the support offered during the process, and how the program aligns with teachers' needs and classroom realities.

This highlights the importance of integrating adequate duration with high-quality, engaging-- a balance reflected in the EDSTEMA model, where sustained engagement is coupled with coherent, content-rich sessions and collaborative reflection to maximize teacher learning and practice.

In sum, the research clearly demonstrates that time is a necessary but not sufficient condition for effective PD. Duration enhances PD impact only when accompanied by quality instruction, relevance, and opportunities for feedback and collaboration. The sustained structure of EDSTEMA reflects this principle, providing teachers with the necessary time and support to deepen SCK, strengthen STE, and ultimately enhance SAA. This alignment underscores the methodological justification for measuring changes in teacher outcomes across an extended implementation period, rather than through short-term interventions.

Collective Participation

Another important characteristic is *collective participation*, which emphasizes the value of teachers working together in teams. Some researchers advocate for the practice of having groups of teachers from the same subject area or grade level engage in professional development (PD) together (Ball, 1996; Knapp, 1997). Supporters of this practice argue that teachers who participate together can serve as a support network for one another, engaging in substantive discourse about their teaching practices and classroom challenges. This type of collaborative learning allows teachers to exchange ideas, share experiences, and solve problems collectively, creating a sense of community that strengthens their teaching abilities.

Research supports the idea that collective participation in PD leads to more substantial changes in teacher practices. For example, Garet et al. (1999; 2001) found that PD programs incorporating collective participation positively influenced teacher practices. When teachers collaborate within their subject areas or grade levels, they are more likely to engage in ongoing, deep conversations about their content and pedagogy.

This type of professional learning (PL) fosters a culture of continuous improvement, which is often lacking in traditional, isolated PD settings. This is directly relevant to the present study because Every Day STEM for All (EDSTEMA) was intentionally structured to include collaborative teams of teachers who engaged in shared planning, content discussions, and reflection, which are key mechanisms through which collective participation fosters professional growth.

Collaborative PD can help bridge the gap between theory and practice. When teachers work together on the same curriculum or teaching strategies, they can directly apply what they have learned in their own classrooms. This practical, application-oriented approach has been shown to enhance the relevance and effectiveness of PD, making it more likely that teachers will implement new strategies successfully (Desimone, 2011; Penuel et al., 2007).

For example, Penuel et al. (2007) found that teachers involved in content-focused, group-based PD were more successful at integrating new teaching strategies into their classrooms than those who participated in more traditional, individual PD sessions. Within the EDSTEMA framework, collective participation was embedded through collaborative inquiry projects and shared reflection sessions that allowed teachers to test and refine new practices together, reflecting a deliberate design choice informed by this research base.

Additionally, collective PD fosters a sense of accountability among teachers. When groups of teachers participate in PD, they can observe one another's practices, provide constructive feedback, and hold each other accountable for applying what they learn. This peer support system was identified as a key factor in sustaining the impact of PD over time (Guskey, 2002). Teachers who participate in collaborative PD are more likely to remain committed to implementing new strategies, knowing that they have the support of their colleagues. This aligns with the EDSTEMA model's inclusion of peer feedback and follow-up discussions, which created cycles of shared accountability and encouraged teachers to reflect openly on both successes and challenges.

Collective participation also contributes to teachers' professional satisfaction and empowerment. Teachers who regularly discuss their instructional experiences with their colleagues report feeling more supported and valued, leading to increased morale and commitment to continuous improvement. These benefits extend beyond formal PD sessions. Collaborative PD often sparks the formation of informal *professional learning networks* (PLNs), where teachers share resources, advice, and encouragement (Hourd, 1997). In this study, participants frequently continued their collaboration beyond the formal EDSTEMA sessions, illustrating the sustainability and strength that collective participation can generate within school communities.

Despite its benefits, implementing collective PD presents several challenges. Factors such as time constraints, scheduling conflicts, and resistance from teachers unaccustomed to collaboration can limit the effectiveness of group-based PD initiatives (Hourd, 1997).

However, research indicates that overcoming these obstacles is worthwhile, as the benefits of collective PD far outweigh the initial challenges. Allowing time for teachers to collaborate and reflect together is an essential element of effective PD programs, particularly in schools focused on fostering long-term professional growth. EDSTEMA addressed these potential barriers by structuring its PD sessions during the academic year and embedding collaboration time within each module, ensuring teachers could engage meaningfully without adding unsustainable demands to their workload.

In sum, collective participation is a cornerstone of effective PD because it builds professional community, strengthens instructional coherence, and supports the long-term application of new practices. When teachers learn together, they not only enhance their individual expertise but also contribute to a culture of shared accountability and collective efficacy. This principle is central to the design and methodological justification of EDSTEMA: its cohort-based model leverages collective participation as a means for deepening teacher science content knowledge (SCK, enhancing science teaching self-efficacy (STE), and improving student academic achievement (SAA). By engaging teachers within a collaborative, reflective community, the program embodied the evidence-based understanding that sustained, collective engagement drives lasting instructional improvement.

Professional Development Type

Type refers to how the professional development (PD) program is structured and deployed. Numerous studies have found that innovative forms of PD, such as study groups, professional learning communities (PLCs), or professional learning networks (PLNs), are more effective at bringing about teacher change than traditional PD formats like workshops or lectures (Ball, 1996; Darling-Hammond, 1995, 1996; Little, 1993). These alternative PD models often emphasize collaboration, reflection, and sustained engagement over time, providing teachers with opportunities to interact with colleagues and learn from one another in meaningful ways.

Unlike traditional one-off workshops or lectures—which often deliver content in a passive manner—innovative PD models foster active learning environments where teachers can engage in problem-solving, share experiences, and build on each other's insights (Vescio et al., 2008). This allows teachers to work together to deepen their content understanding, improve their teaching strategies, and learn from each other's experiences.

For example, study groups provide a structured way for teachers to engage in regular, focused discussions about instructional practices and student outcomes, fostering an environment of mutual support and professional growth (Hourd, 1997). These groups encourage collaboration, enabling teachers to problem-solve together and exchange strategies to address the unique needs of their students. As such, study groups contribute not only to individual teacher growth, but also to building a community of support among colleagues.

The collaborative nature of these groups fosters a culture of continuous improvement and helps create an atmosphere of trust and respect, both of which are essential for long-term professional growth (Borko, 2004). This emphasis on collaborative, ongoing structures directly aligns with the current study, as Every Day STEM for All (EDSTEMA) was intentionally designed as a hybrid PD model incorporating elements of both traditional workshops and sustained collaborative sessions to balance accessibility with depth and reflection.

However, while innovative PD models like study groups, PLCs, and PLNs are promising, the relationship between PD type and changes in teacher practice is complex. For instance, Garet et al. (1999, 2001) did not find a direct relationship between PD type and improvements in teacher practice. Instead, they argued that PD type can indirectly influence teacher practice by affecting other factors, such as the duration of the program, the level of teacher engagement, and the alignment between PD content and teachers' specific needs. According to their findings, PD that is sustained over time, actively engages teachers, and has direct practical applicability to teachers' classrooms is more likely to result in lasting change (Garet et al., 2001).

This suggests that the quality and structure of PD, rather than its format, may be the most important determinants of its success. This distinction is reflected in the design of the EDSTEMA program, which blended traditional face-to-face sessions with sustained collaborative components and classroom-based application cycles to maximize both accessibility and depth.

In their critical analysis of PD effectiveness, Guskey and Yoon (2009) questioned whether innovative features are essential for PD to be effective. Drawing on the meta-analysis by Yoon et al. (2007), they noted that all of the programs—despite using traditional formats such as workshops or summer institutes—showed a positive relationship between PD and student academic achievement (SAA). Specifically, they found that this positive relationship held true across all studies, regardless of the PD format.

Guskey and Yoon (2009) argued that workshops should not be dismissed as the "poster child of ineffective practice" they are often perceived to be. Instead, they emphasized that the success of PD depends on its design, duration, content, and how well it addresses the individual needs of teachers. This is particularly relevant to the current study because EDSTEMA embedded traditional workshop experiences within a coherent sequence of collaborative reflection, follow-up activities, and classroom application, as these elements were shown in the literature to sustain impact beyond initial delivery.

This body of evidence helps explain why Desimone (2009) did not include "type" as one of her five critical features of effective PD. While Garet et al. (1999, 2001) had identified type as a structural feature alongside duration and collective participation, subsequent research showed that type itself had little direct effect on teacher practice. Rather, its influence was indirect, depending on whether the PD incorporated more substantive features such as active learning, content focus, and coherence. Building on this evidence, Desimone streamlined the framework to focus on the five features most consistently associated with meaningful teacher and student outcomes: content focus, active learning, coherence, sustained duration, and collective participation.

This raises a critical question for the design of PD programs—Which is more important in determining student outcomes: the format of the activities or the quality of content? While both factors likely play a role in the success of PD, the key is ensuring that PD programs are not isolated events but rather part of a continuous process that provides ongoing support to teachers in meaningful ways.

Effective PD can take many forms—whether traditional or innovative—as long as it encourages active participation, promotes peer collaboration, and aligns with teachers' specific needs. In this sense, EDSTEMA's blended structure reflects the literature's guidance by combining accessible workshop-style sessions with inquiry-based collaboration and classroom follow-up, ensuring that format serves as a mechanism for sustained, content-rich professional learning rather than an end in itself.

Summary of Professional Development Literature

In sum, the literature on professional development highlights several features that consistently contribute to its effectiveness, including content focus, active learning, coherence, sustained duration, and collective participation.

Research has also shown that while the format or type of professional development (PD) can shape opportunities for learning, its impact is mediated by the quality of design and alignment to teachers' needs (Garet et al., 1999, 2001; Desimone, 2009; Guskey & Yoon, 2009).

Collectively, these findings emphasize that effective PD is not defined by novelty or format but by its ability to engage teachers in meaningful, sustained, and contextually relevant learning experiences. Building on this foundation, the next section examines research on science-focused PD, which presents unique challenges and opportunities related to teacher content knowledge, pedagogy, and student outcomes.

Science Teacher Professional Development

Gottfried (1985) emphasized the need for professional development (PD) that focuses specifically on science at the elementary level, noting both the unique pedagogy of the subject and teachers' lack of confidence in teaching it. Addressing these challenges is critical, as teachers' comfort and skill in science instruction directly influence student engagement and achievement.

Research indicates that PD programs incorporating the characteristics identified in Desimone's (2009) framework for effective PD—sustained duration, content focus, collective participation, active learning, and coherence—are particularly effective in strengthening teacher content knowledge, boosting self-efficacy, and improving classroom practices. These improvements, in turn, have been shown to lead to measurable gains in student achievement.

The following sections review a selection of studies that illustrate these outcomes and provide a foundation for understanding how the Every Day STEM for All (EDSTEMA) program builds upon this body of evidence.

Content Understanding

Effective professional development (PD) plays a crucial role in enhancing science content understanding (SCK), which is essential for improving teaching practices, and ultimately, student learning outcomes. For science educators, a deep understanding of subject matter is necessary not only to convey accurate information but also to foster scientific literacy and critical thinking skills among students. As science standards and curricula continue to evolve and increase in complexity, PD must prioritize helping teachers strengthen their SCK while also equipping them with the pedagogical strategies needed to make this knowledge accessible to learners. When teachers are supported in both content and pedagogy, they are better prepared to create engaging, accurate, and meaningful science learning experiences that promote long-term student success.

In recent years, a body of research has examined the impact of PD on science teachers' content knowledge (CK), highlighting a range of promising approaches and strategies. Evidence from these studies demonstrates that PD programs integrating content-focused instruction, collaborative learning opportunities, and sustained engagement over time can produce substantial improvements in both teaching practices and student outcomes.

This section reviews key studies that specifically investigate how PD enhances teachers' understanding of science concepts, providing concrete examples of successful program designs. By synthesizing findings across these studies, the review identifies the essential features of PD that most effectively deepen teachers' science knowledge and offers insights into how future PD initiatives can be structured to better support educators.

Project ICAN

Project ICAN was a professional development (PD) program sponsored by the National Science Foundation (NSF) that examined both student and teacher understandings of the nature of science (NOS) and scientific inquiry (Lederman & Lederman, 2012). While the program primarily focused on secondary science teachers, a number of elementary teachers also participated, which broadened its impact across grade levels. Over the course of its five-year duration, the program engaged 238 teachers and approximately 23,500 students in the Chicago area, making it one of the largest initiatives of its kind.

By centering on NOS and inquiry, Project ICAN demonstrated the potential of long-term, content-focused PD to influence teacher practices and improve student learning on a large scale.

Teachers who participated in Project ICAN engaged in a variety of professional learning experiences designed to strengthen their understanding of NOS. They attended full-day workshops once a month for an entire year, with each session focused on hands-on NOS activities. In addition, participants took part in a 10-day summer institute (SI), where they collaborated on curriculum development to support classroom implementation. Each teacher was paired with a researcher at the Illinois Institute of Technology for an internship, which provided the opportunity to observe authentic scientific investigations firsthand. This combination of workshops, SIs, and research internships exemplified a multi-faceted PD model that connected theory, practice, and authentic scientific experiences.

The researchers found that most participants significantly increased their understanding of NOS and inquiry through participation in the program. In addition to strengthening their own knowledge, teachers also improved their instructional methods for teaching these concepts, integrating new strategies into their classrooms more effectively. By the fourth and fifth years of the program, over 60% of students demonstrated satisfactory understanding of more than 80% of the key aspects of NOS and inquiry, reflecting the positive impact of teacher learning on student outcomes. These findings underscore the value of sustained, content-rich PD experiences like Project ICAN in fostering both teacher growth and measurable improvements in student learning.

The outcomes of Project ICAN highlight important implications for teacher PD, particularly in linking content understanding with pedagogical effectiveness. The program demonstrated that sustained, hands-on PD—through monthly workshops, SIs, and research internships—deepened teachers' understanding of NOS and inquiry while enhancing their instructional methods.

These outcomes underscore the value of PD that incorporates duration (multi-year engagement), content focus (NOS and inquiry), and active learning (hands-on workshops and internships). Moreover, the program promoted collective participation by engaging teachers in ongoing collaborative work, and it exhibited coherence by aligning NOS and inquiry with state and district science goals. Together, these features illustrate how Desimone's (2009) framework provides a lens for understanding the program's success.

Dailey's STEM Initiative

A STEM-focused professional development (PD) was established in four elementary schools to support science teaching in grades two to five (Dailey, 2013). The goal of the initiative was to strengthen teacher content knowledge (CK) and pedagogical content knowledge (PCK) by implementing PD focused on STEM education. The PD developers selected a STEM approach rather than concentrating solely on science because they believed that STEM PD would appeal to a broader range of teachers. Many elementary educators might not be as interested in science-specific PD, but a STEM focus was considered more engaging and relevant, with the potential to integrate science, technology, engineering, and mathematics into daily instruction. By framing the initiative in terms of STEM, the program also aimed to encourage interdisciplinary connections and foster student interest in these fields at an earlier age.

The study employed an experimental design, with teachers randomly assigned to either the treatment (TG) or comparison group (CG). The TG participated in a two-week-long summer institute (SI), and received job-embedded peer coaching support throughout the year. The researcher found that changes in teachers' perceptions of elementary science teaching occurred after one year, or 60 hours of PD. However, significant changes in teachers' science process skills were observed only after two years, or 120 hours of PD. Notably, these changes in teachers' teaching perceptions, concerns, and science process skills held constant after PD supports had been removed for an entire year.

Dailey's study of a STEM-focused PD initiative in four elementary schools showed that recruiting teachers through a broader STEM lens encouraged participation from educators who might have resisted science-specific PD. Teachers in the TG participated in SIs and received peer coaching, which provided both intensive learning and ongoing job-embedded support. After 60 hours of PD, teachers reported changes in their perceptions of science teaching, and after 120 hours, they demonstrated measurable improvements in science process skills.

Importantly, these changes persisted even after PD supports were removed for a year, indicating lasting effects. The structure of this program reflects several features of Desimone's (2009) framework: duration (two years of sustained PD), active learning (hands-on STEM activities), and coherence (alignment with STEM education priorities and district goals). The program also fostered collective participation through peer coaching and collaboration. These connections demonstrate how well-designed, framework-aligned PD can produce both immediate and enduring changes in teacher practice.

Zwiep's Content-Rich Learning Experience

In Zwiep's study on content-rich learning experiences, teachers first participated in a week-long summer institute (SI) followed by a series of on-site meetings within their districts (Zwiep, 2012). During the SIs, teachers were organized into groups based on grade level or subject area, which allowed them to collaborate with colleagues who faced similar instructional challenges.

The primary goal of the SI was to strengthen teacher content knowledge (CK) and pedagogical content knowledge (PCK) by addressing both instructional practices and teacher beliefs. In particular, the program emphasized the use of effective pedagogical techniques for reaching all students and included activities where teachers analyzed student work to assess conceptual understanding and monitor student growth. By combining collaborative learning, targeted pedagogy, and analysis of student outcomes, the program aimed to foster meaningful improvements in both teaching practice and student achievement.

Researchers used a combination of quantitative tests and qualitative reflections to evaluate the impact of the program. Their findings revealed that teachers' CK improved significantly after participating in the professional development (PD). In addition, participants demonstrated growth in several key areas, including their perceptions of the nature of each discipline, their attitudes toward learning new content, and their views about students' understandings. Teachers also reported greater confidence in their own content knowledge as well as in their ability to teach science effectively. Collectively, these results suggest that the program not only enhanced teachers' knowledge but also positively influenced their beliefs, attitudes, and instructional confidence, creating a more comprehensive improvement in teacher capacity.

Zwiep (2012) investigated the impact of a PD program that combined a week-long SI with district-based follow-up meetings. Teachers were grouped by grade level or subject area, allowing them to collaborate with peers who shared similar instructional challenges. The SI emphasized enhancing both CK and PCK by engaging teachers in lesson analysis, curriculum development, and the examination of student work for conceptual understanding. Quantitative and qualitative findings indicated significant improvements in teachers' CK and confidence, along with more positive attitudes toward student learning.

This program illustrates multiple aspects of Desimone's (2009) framework: duration (summer institute plus follow-up meetings), content focus (deepened SCK and PCK), active learning (lesson planning and analysis), and collective participation (teacher collaboration within grade-level groups). Coherence was achieved by ensuring the PD aligned with district instructional goals and classroom realities. Together, these features highlight the importance of designing PD that is both intensive and sustained to achieve meaningful teacher and student outcomes.

Science Teaching Self-Efficacy

Professional development (PD) programs aimed at enhancing science teaching have consistently been shown to positively influence science teaching self-efficacy (STE), which is a critical factor in teaching effectiveness, due to its ability to shape teachers' motivation, instructional strategies, and overall approach to classroom practice.

Research demonstrates that PD programs offering sustained, content-focused learning opportunities can strengthen teachers' confidence in their ability to teach complex science concepts and implement inquiry-based practices. By fostering greater confidence, PD not only improves teachers' instructional skills but also encourages them to take pedagogical risks, explore new teaching methods, and create more engaging learning experiences for their students.

Penuel et al. (2007) examined the effects of a PD program specifically designed for science teachers, with a focus on inquiry-based science instruction. The study found that teachers who participated in the program, which involved sustained professional learning (PL) experiences and active learning strategies, reported significant improvements in their teaching self-efficacy (TSE). Teachers expressed greater confidence not only in their ability to teach complex science content but also in their capacity to engage students in inquiry-based scientific investigations. This increase in TSE stemmed from the program's emphasis on providing teachers not only with the knowledge and skills required to teach science effectively, but also with opportunities to actively practice and apply these skills in a supportive, collaborative environment.

Desimone et al. (2002) supported this finding, emphasizing that PD programs that are sustained over time, content-focused, and include active learning components significantly enhance teachers' confidence in their teaching abilities. Their research highlighted the importance of PD programs that go beyond one-off workshops, showing that ongoing learning opportunities allow teachers to deepen their content knowledge (CK) and refine their pedagogical practices.

Furthermore, PD programs that foster collaboration—by enabling teachers to share experiences and insights—also contribute to increased TSE. Collaboration provided teachers with a sense of community and shared purpose, reinforcing their belief in their ability to improve their teaching practices and overcome challenges in the classroom.

Both studies emphasize the importance of PD programs that not only focus on content but also incorporate strategies that actively engage participants. When teachers are given opportunities to apply new knowledge, collaborate with peers, and reflect on their practice, they are more likely to build confidence and strengthen their TSE. This growth in confidence equips teachers with both the tools and the belief in their ability to teach science effectively. As a result, they are better prepared to engage students in meaningful learning experiences that foster deeper understanding of scientific concepts. Ultimately, PD that blends content focus with active engagement creates a cycle of teacher growth that directly benefits student achievement.

These studies suggest that PD programs that focus on both CK and teaching practices, while providing opportunities for active learning and collaboration, are essential in fostering greater TSE in science teachers. Enhanced self-efficacy leads to more effective teaching strategies, which ultimately enhance student engagement and achievement in science.

As TSE is a key factor in determining teaching effectiveness, PD programs that allow teachers to improve both their science content understanding (SCK) and confidence should be prioritized in efforts to improve science education. The following section describes a PD program shown to improve participants' STE.

Kenan Fellows Program

Powell-Moman and Brown-Schild (2011) examined the effects of a two-year professional development (PD) on science teachers' teaching self-efficacy (TSE) and use of inquiry-based instruction. The Kenan Fellows program was designed to help teachers build confidence in teaching science and promote the use of inquiry-based, student-centered teaching strategies. The PD took place over two years and focused on fostering inquiry-based instruction, which emphasizes active student engagement in the scientific process. The program provided teachers with opportunities to engage in hands-on science activities, participate in workshops, and collaborate with peers and other stakeholders. This structure aimed to develop both teachers' content knowledge (CK) and pedagogical skills, ensuring they were equipped to implement inquiry-based practices in their classrooms. Teachers also received ongoing support throughout the program, including opportunities for reflection, discussion, and peer collaboration. This collaborative environment was crucial in helping teachers feel supported and encouraged as they piloted new strategies.

The study found that the two-year PD program had a significant positive impact on teachers' TSE. Teachers reported increased confidence in their ability to teach science effectively, particularly in utilizing inquiry-based approaches. They felt more capable of designing and implementing lessons that fostered hands-on, student-centered learning activities. Additionally, participants' use of inquiry-based instruction increased after participating in the program. The researchers attributed this to the sustained nature of the PD, which allowed teachers to gradually integrate new techniques and reflect on their practices in a supportive, collaborative environment.

The study also emphasized the importance of sustained PD, noting that a two-year commitment enabled teachers to build upon their learning over time. This extended duration provided opportunities for teachers to revisit concepts, practice new strategies, and refine their teaching practices as they gained experience. As a result, teachers were able to develop a deeper and more confident understanding of inquiry-based methods that translated into meaningful classroom application. The ongoing support from peers and program facilitators further helped maintain momentum, ensuring that teachers stayed engaged and had the guidance needed to troubleshoot challenges. Collectively, these factors highlight how long-term, well-structured PD can lead to lasting improvements in both teacher practice and student learning.

The findings from this study suggested that long-term, inquiry-based PD can significantly strengthen teachers' TSE. By offering extended opportunities for active learning, collaboration, and reflection, such programs allow teachers to gradually build confidence in their ability to teach science and to implement more effective, student-centered instruction. Sustained PD experiences that emphasize both content knowledge (CK) and pedagogy are especially valuable, as they provide teachers not only with the knowledge and strategies needed to improve their practice but also with the support structures necessary for continuous growth.

Ultimately, these findings reinforced the argument that well-designed PD can serve as a powerful lever for enhancing teacher effectiveness and improving student learning in science. In this way, the study exemplified the critical features of effective PD outlined in Desimone's (2009) framework—namely, duration, content focus, active learning, coherence, and collective participation—which together create the conditions for meaningful and lasting change.

Teaching Practices

Building on the research reviewed in the previous section, studies have consistently shown that well-designed professional development (PD) programs play a pivotal role in equipping educators with the pedagogical strategies, content knowledge (CK), and instructional tools necessary to improve student outcomes. In science education, PD is particularly critical because it enables teachers to stay current with evolving standards, deepen their conceptual understanding, and adopt effective strategies for fostering inquiry-based learning.

The sections that follow examine various PD models and assess their effectiveness in enhancing science teaching practices. By exploring the connection between PD and teaching practice, this review underscores how intentionally structured professional learning (PL) opportunities can not only improve classroom instruction but also contribute to the broader goal of advancing science literacy among students.

Physical Science Inquiry Academy

In a study by Drits (2011), elementary teachers participated in a two-year professional development (PD) program designed to enhance their science teaching practices. The program began with a summer institute (SI) that introduced teachers to the 5E model of scientific inquiry, a widely adopted instructional framework for structuring science lessons to promote active learning and student engagement. This phase was followed by 10 monthly follow-up sessions, during which teachers were able to deepen their understanding of the 5E model and incorporate it into their own teaching practices. In addition, teachers received classroom support from an elementary science specialist who also conducted three classroom observations annually. This combination of structured training, continuous follow-up, and individualized support created a sustained professional learning environment that encouraged teachers to apply new strategies with confidence and consistency.

To assess the impact of the program, teachers completed content assessments and self-efficacy inventories at various times throughout the two years. The data collected revealed significant improvements in teachers' science content understanding (SCK) and teaching self-efficacy (TSE), suggesting that the program was effective in enhancing their confidence and expertise in science teaching. Notably, however, the study also found that—while there were significant changes in teacher practices—these changes were most pronounced during the first semester of the two-year program, after which the improvements appeared to plateau.

These findings highlighted the potential for immediate improvements in teacher practices following an intensive PD program, although they also suggested that sustained, long-term change might require ongoing support and follow-up beyond the initial phases of the program. Overall, the study underscored the value of PD programs that combine content knowledge (CK), pedagogical content knowledge (PCK), and classroom support, while also suggesting that the longevity of PD effects may depend on the nature and frequency of ongoing professional learning (PL) opportunities.

Investigating Our Blue Planet

The purpose of Tinnin's (2000) study was to determine the effectiveness of 30 days of science professional development (PD) delivered over 24 months at a single school site. The extended program aimed to deepen teachers' understanding of science instruction by integrating thematic science strands, guided inquiry activities based on the 5E model, and an emphasis on the interdisciplinary nature of science process skills. To assess the program's impact, the researcher evaluated 15 participating teachers in the areas of teaching self-efficacy (TSE), attitudes toward science, instructional strategies, and content knowledge. This approach provided a comprehensive look at how sustained, inquiry-based PD fostered both teacher growth and more effective science teaching practices.

The findings of the study indicated that the PD was successful in improving several aspects of the teachers' professional growth. Notably, the teachers demonstrated significant gains in content knowledge (CK) and a deeper understanding of science concepts. Teachers also showed substantial improvements in TSE, reflecting increased confidence in their ability to teach science effectively. They further reported meaningful shifts in their classroom practices, adopting more inquiry-based strategies and integrating interdisciplinary approaches into their instruction. Taken together, these results underscore the potential of well-designed PD programs that purposefully integrate CK and pedagogical content knowledge (PCK) to promote lasting teacher growth and improved classroom practice.

Overall, the study highlighted the effectiveness of sustained PD efforts, particularly those that combined theoretical and practical elements to foster teacher growth. The findings demonstrated that long-term, structured PD significantly improved multiple dimensions of teacher development, including CK, attitudes toward science, and instructional approaches. These improvements extended beyond technical skills, influencing teachers' confidence and willingness to embrace more inquiry-based and interdisciplinary methods. Collectively, the results provided compelling evidence for the value of comprehensive, ongoing PD opportunities that support both teacher confidence and competence in science instruction, ultimately contributing to stronger student learning experiences.

Physics and Everyday Thinking

This study highlighted the importance of measuring different aspects of teacher learning resulting from a professional development (PD) program (Harlow, 2013). In this particular study, teachers participated in a PD focused on content and pedagogy related to elementary physics, specifically electricity and magnetism. Through pre- and post-classroom observations, the researchers found that each of the three teachers showed growth in different ways as a result of the workshop, highlighting the individualized impact of PD. While all three teachers demonstrated growth in their content knowledge (CK) of electricity and magnetism, each teacher developed specific skills: one enhanced her understanding of the predictive power of scientific models, another improved her ability to encourage students to formulate testable questions, and a third strengthened her capacity to design more open inquiry-based lessons.

These differentiated outcomes underscore the individualized impact that PD can have on teaching practices, as teachers may respond and grow in unique ways depending on their prior knowledge and classroom contexts. The results reinforce the idea that sustained and targeted PD programs are essential for promoting comprehensive teacher growth, allowing teachers to deepen their content knowledge while also expanding their instructional repertoire.

Importantly, the findings make a compelling case for the need for PD programs that provide not only content-specific knowledge but also adequate time and structured support for teachers to refine and diversify their pedagogical approaches. Taken together, this evidence suggests that well-designed PD must balance subject matter expertise with opportunities for reflection, practice, and collaboration if it is to generate meaningful and lasting improvements in both teaching and student learning.

Student Achievement

This section explores the relationship between professional development (PD) for science teachers and increases in student academic achievement (SAA). A growing body of research indicates that well-designed PD programs can significantly strengthen teaching quality, which, in turn, has a direct and positive effect on student learning outcomes in science. By reviewing studies that link teacher growth to student performance, this section will highlight the mechanisms—such as enhanced content knowledge (CK), improved instructional strategies, and increased teacher self-efficacy (TSE)—through which PD exerts its influence. Understanding these connections is critical for designing PD initiatives that go beyond supporting teacher development alone and instead translate into measurable gains in student learning. Ultimately, this analysis underscores the dual role of PD as both a tool for teacher professional growth and a strategy for elevating student achievement in science education.

Transformative Professional Development

In this case study, one school received professional development (PD) over two years based on the principles for effective PD outlined by Desimone (2009), while another did not (Johnson & Fargo, 2014). The goal of the PD was to transform science teaching and learning through the enhancement of teaching effectiveness. The PD program required participants to attend a two-week summer workshop, eight full-day workshops during the school year, and monthly grade-level support sessions. The workshops were designed to align with state science standards. The researchers reported that the school that received the PD intervention raised the percentage of students scoring proficient on the state test from 25% to 67%. These findings indicate a clear connection between structured, sustained PD and student academic achievement (SAA). This study highlighted the potential of thoughtfully designed PD programs to positively influence both teaching practices and student outcomes, especially when they are aligned with state standards and continuously supported.

A Teachers Workshop in Utilizing the Learning Cycle in Elementary School Science

Marek & Methven (1991) investigated the relationships between teacher attitudes and the implementation of the learning cycle model, and the impact of such instruction on elementary students' reasoning and language skills. This study involved 25 participants from seven states, all of whom took part in a four-week, 20-day summer professional development (PD) workshop. The workshop was designed to engage participants in science content through the *learning cycle model of inquiry*, a student-centered approach that emphasizes hands-on exploration, content development, and reflective thinking. During the workshop, teachers experienced several learning cycle lessons as learners and took turns teaching each other using pre-written learning cycles. To further reinforce the model, participants were provided with a book of learning cycle lessons and were tasked with teaching these lessons in their own classrooms.

To assess the impact of the PD experience, participants completed questionnaires that examined their science teaching attitudes and practices. Throughout the following year, the teachers were observed as they implemented the learning cycle model in their classrooms. The study revealed significant gains in reasoning abilities and language skills in the classes where the learning cycle model was implemented. This suggests that the PD program, which focused on deepening teachers' understanding of inquiry-based teaching methods, not only influenced teacher practices but also contributed to meaningful gains in student learning outcomes. Overall, the findings underscore the potential of PD programs that equip teachers with inquiry-based teaching models to strengthen teacher attitudes and improve student academic achievement (SAA), particularly in areas such as science reasoning and communication.

Three Professional Development Models

In response to a meta-analysis by Yoon et al. (2007), which identified a need for more rigorous evidence on the impact of professional development (PD) on teaching and student outcomes, Heller et al. (2012) designed a study to expand the research base for PD and meet higher standards of evidence. The researchers employed an experimental design to examine the differential effects of three types of PD interventions. Each intervention was structured to align with the key recommendations for effective PD outlined by Desimone (2009), which include content focus, active learning, coherence, substantial duration, and collective participation.

The first intervention, *Teaching Cases*, engaged teachers in discussions of narrative classroom cases integrating science content, student thinking, and instruction around content-based practical dilemmas. These cases featured real-world scenarios, including descriptions of instructional activities, student work samples, student-teacher dialogues, and reflections on teacher actions and behaviors. The second intervention, *Looking at Student Work*, asked teachers to analyze their own students' work after teaching a unit on electric circuits. The third intervention, *Metacognitive Analysis*, engaged teachers in reflective discussions about their own learning experiences.

The study found that all three PD interventions led to significant improvements in teacher content knowledge (CK), with large gains in teacher performance on content assessments. Moreover, the gains in teacher science content understanding (SCK) were evident not only during the treatment year but also translated into significant improvements in student academic achievement (SAA)—measured by student content test scores—in the classrooms of participating teachers, even one year later. These findings underscore the potential that PD of moderate duration can produce lasting and meaningful changes in both teacher practices and student learning outcomes in elementary science education. The study thus provides strong evidence that well-designed, content-focused PD can have a substantial and enduring impact on teaching quality and SAA, further supporting the importance of investing in high-quality PD for educators.

Science Professional Development Summary

In summary, the body of research on science-focused professional development demonstrates that programs grounded in Desimone's (2009) framework—emphasizing content focus, active learning, coherence, sustained duration, and collective participation—consistently improve teacher knowledge, strengthen self-efficacy, and enhance instructional practice. These improvements have been shown to translate into measurable gains in student achievement, underscoring the potential of well-designed professional development (PD) to address longstanding challenges in science education. Building on these findings, large-scale initiatives such as the Mathematics and Science Partnership (MSP) grants have sought to institutionalize these features within district- and state-level reforms. The next section reviews MSP-funded programs, highlighting how they operationalized the critical features of effective PD at scale and how their outcomes informed the design and implementation of Every Day STEM for All (EDSTEMA).

Math and Science Partnership Grants

A crucial component of national efforts to improve K–12 education in the United States was the Mathematics and Science Partnership (MSP) program, a class of professional development (PD) initiatives designed to strengthen mathematics and science instruction. Research (e.g., Bolyard & Moyer-Packenham, 2008) has consistently shown that teacher quality is one of the most significant school-based factors influencing student academic achievement (SAA). In response to this evidence, the federal government—through agencies such as the National Science Foundation (NSF) and the Department of Education—developed and funded the MSP program to improve teaching and learning in these subject areas. Launched in 2002 under Title IIB of the No Child Left Behind (NCLB) Act (NSF, 2010) and discontinued in 2016, the MSP program aimed to ensure a stronger pipeline of highly qualified teachers in mathematics and science, addressing a critical need in American education.

MSP grants were competitive, merit-based awards given to collaborative teams composed of institutions of higher education, local K–12 school systems, and other supporting organizations. The overarching goal of all MSPs was not only to enhance teacher effectiveness but also to improve SAA in mathematics and science. A very important component of the MSP mission was to significantly reduce achievement gaps in math and science among diverse student populations. By focusing on improving the instructional practices of teachers, the program sought to ensure that all PreK–12 students, regardless of background, received a high-quality education in math and science.

To achieve these ambitious outcomes, the MSP program invested millions of dollars in the developing, implementing, and sustaining innovative PD programs. These initiatives were designed to provide teachers with the content knowledge (CK), pedagogical strategies, and ongoing support necessary to transform their classroom practices and improve student learning. Over the course of 15 years, MSP grants funded a variety of PD programs, ranging from elementary science-focused workshops to high school-level curriculum development initiatives.

The sections that follow describe two MSP programs specifically targeting elementary science PD, highlighting their design, implementation, and impact on both teacher practices and SAA. Importantly, the lessons learned from these MSP-funded initiatives provided the foundation and rationale for the creation of the Every Day STEM for All (EDSTEMA) program, which sought to address persistent challenges in elementary science education through a sustained, coherent model of professional learning.

Milwaukee Urban Systemic Initiative Project

In this study, Posnanski (2002) intentionally designed a professional development (PD) program aimed at improving elementary teachers' science teaching self-efficacy (STE). The research was conducted as part of the Milwaukee Urban Systemic Initiative Project, a multi-year Mathematics and Science Partnership (MSP) initiative. A total of 43 elementary teachers participated in the project, which involved PD sessions lasting 3 to 4 hr per week over the course of 32 weeks. Teachers met in school-based teams of two to four members. The focus of these meetings was on addressing practical science teaching issues and providing opportunities for teachers to enhance their instructional practices.

The PD sessions combined collaborative discussions on science teaching issues, model lessons using novel curricula, and demonstrations of effective teaching strategies led by experienced educators. During these sessions, participants were encouraged to engage in hands-on activities, experiment with alternative teaching strategies, and video record themselves while teaching to capture their instructional practices. In addition, teachers received feedback from both grant personnel and their peers, which provided opportunities for reflection and professional growth. To further support their learning, teachers were also asked to maintain a project portfolio and regularly reflect on their experiences through journaling.

The results of the study indicated that this PD program had a positive impact on teachers' beliefs about their ability to teach science effectively. The vast majority of participants reported that the program positively influenced their ability to use inquiry-based activities and that they felt more comfortable using hands-on materials in their science classrooms. Additionally, all participants agreed that the PD helped them develop a deeper understanding of various teaching approaches and gave them a stronger grasp of how to teach science more effectively in their classrooms. Teachers also reported greater clarity regarding the science curriculum, including its outcomes, standards, and methods of facilitation.

Moreover, the study revealed that the PD program had lasting implications for teachers' intentions to improve their practice. Nearly all participants responded that they intended to provide more science learning opportunities for their students, signaling their increased commitment to science education following their participation in the PD.

Partnership for Effective Science Teaching and Learning

This study by Hauck (2012) sought to investigate the extent to which professional development (PD) influenced elementary classroom instruction, particularly in science education. The researcher utilized existing data from 22 randomly sampled teachers who had participated in a three-year PD program. The program was designed to enhance teachers' understanding of scientific phenomena and improve their instructional practices. The PD began with a five-day summer seminar aimed at deepening teachers' science content understanding (SCK). Following the seminar, the PD continued throughout the school year through school-based professional development communities (PLCs) facilitated by fellow teachers. These meetings consisted of readings, discussions, lesson ideas, and reflective journaling. Additionally, grade-level sessions were held two to three times during the year, where program facilitators helped teachers align their grade-level standards with their teaching resources. Mid-winter institutes provided further opportunities for teachers to engage in inquiry activities, hone science practices, and explore the nature of science (NOS).

This study's data consisted of classroom observations conducted with a program observation protocol that assessed five key areas of instruction: science talk and argumentation, investigation, modeling, alignment of science content to standards, and addressing science misconceptions. The findings indicated that participation in the program had a significant positive impact on classroom instruction, suggesting that teachers were better able to engage students in authentic science practices. Notably, teaching practices showed marked improvements in the areas of substantive discourse and argumentation, inquiry-based investigation, modeling scientific concepts, alignment of instruction with standards, and the correction of common student misconceptions. These outcomes highlight the potential of well-structured PD programs to foster meaningful changes in classroom practice, equipping teachers to create richer and more effective science learning environments for their students.

An important finding of this study was that teachers who had received sustained PD demonstrated greater overall effectiveness in science instruction. Specifically, those who participated in the program for the full three years exhibited greater improvements in their instructional practices compared to teachers who engaged for only one or two years. This finding underscores the critical role of duration in determining the success of PD, as longer-term participation appears to foster deeper understanding, stronger skill development, and more lasting changes in teaching practices. Ultimately, the results suggest that short-term PD may provide initial benefits, but it is sustained, multi-year engagement that leads to the most meaningful and enduring improvements in science instruction.

Math and Science Partnership (MSP) Grant Summary

In summary, the Mathematics and Science Partnership (MSP) program represented a significant federal investment in improving teacher quality in mathematics and science through sustained, collaborative, and content-rich professional development. Evidence from initiatives such as the Milwaukee Urban Systemic Initiative Project (Posnanski, 2002) and the Partnership for Effective Science Teaching and Learning (Hauck, 2012) demonstrated that MSP-funded programs could positively influence teacher beliefs, strengthen instructional practices, and promote inquiry-based approaches to science teaching.

Importantly, these studies highlighted the value of long-term, coherent professional development (PD) models that incorporate active learning, collective participation, and alignment with standards. While the national MSP program ended in 2016, the lessons it generated provided a strong foundation for subsequent efforts to address persistent gaps in elementary science education.

Building on this body of evidence, the next section introduces the Every Day STEM for All (EDSTEMA) program, which was designed to extend and refine these principles within a sustained, district-level model of professional learning.

Summary

This literature review established a clear connection between professional development (PD) and improvements in teacher content knowledge (CK), teaching self-efficacy (TSE), instructional practice, and student academic achievement (SAA). Across numerous studies, well-designed PD has been shown to enhance teacher confidence, strengthen pedagogical strategies, and improve student outcomes. Collectively, these findings underscore the importance of sustained, high-quality PD as a mechanism for advancing both teacher learning and student success in science.

Research consistently identifies five essential characteristics of effective PD: content focus, active learning, coherence, sustained duration, and collective participation (Desimone, 2009). When these features are intentionally integrated, PD becomes more meaningful, sustainable, and impactful. Content-focused PD deepens teachers' subject-matter expertise and aligns with curriculum standards. Active learning allows teachers to apply and reflect on new strategies. Coherence ensures alignment with broader school goals, while collective participation fosters collaboration and shared accountability. Sustained duration provides time for teachers to implement, reflect, and refine new practices, leading to lasting change.

Empirical evidence from elementary science-focused PD initiatives, including Mathematics and Science Partnership (MSP) programs, demonstrates that PD structured around these characteristics yields measurable gains in teacher practice and student performance. Teachers who experience such PD report higher confidence in teaching science and greater ability to engage students in inquiry-based learning, while students show improved conceptual understanding and achievement. These findings reinforce that intentional, research-based PD not only transforms classroom instruction but also contributes to broader educational improvement.

Building upon this foundation, the present study examines how Every Day STEM for All (EDSTEMA) operationalized these five features within its design. As shown in Table 1, EDSTEMA integrates content focus, active learning, coherence, sustained duration, and collective participation into a cohesive PD model aimed at improving teacher knowledge, efficacy, and student outcomes. This alignment provides the conceptual and structural basis for the study's methodological design, described in Chapter 3.

Table 1

Alignment of Desimone's Features of Effective PD with the Design of the EDSTEMA Program

Desimone's (2009) Features of Effective PD	Corresponding Elements within the EDSTEMA Program
Content Focus	EDSTEMA emphasized SCK through hands-on, standards-aligned science lessons. Each module deepened teachers' conceptual understanding, directly linking PD content to classroom curricula.
Active Learning	Teachers engaged in inquiry-based and modeling activities that mirrored authentic classroom practice. Participants tested new strategies, analyzed data, and reflected on student misconceptions to reinforce learning through experience.
Coherence	Program content aligned with OAS, district pacing guides, and existing instructional initiatives. Sessions built sequentially to connect prior knowledge with new skills, ensuring continuity between PD and classroom implementation.
Sustained Duration	EDSTEMA extended over multiple months, combining in-person sessions and follow-ups, and classroom application cycles. This structure allowed teachers to apply new strategies, gather student evidence, and reflect collaboratively in subsequent sessions.
Collective Participation	Cohort-based design encouraged cross-grade collaboration among teachers from partner districts. Participants planned lessons together, exchanged feedback, and built a PLN that fostered sustained support beyond the PD period.

Note. EDSTEMA = Every Day STEM for All; PD = professional development; SCK = science content knowledge; OAS = Oklahoma Academic Standards; PLN = professional learning network

Chapter 3: Methodology

This study examined the effects of the Math and Science Partnership (MSP) grant-funded Every Day STEM for All (EDSTEMA) professional development (PD) program on elementary teachers' science content knowledge (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) in science. A quasi-experimental quantitative design was utilized to investigate these relationships. This chapter describes the methodological approach employed in the study. It begins with a description of the PD program's context and the general characteristics of the teacher participants. It then details the data collection procedures, including the instruments used to measure SCK, STE, and SAA. Finally, the chapter explains the data analysis methods used to evaluate the program's effects.

Research Design

This study employed a quasi-experimental pretest–posttest design with nonequivalent groups to examine the research questions. Quasi-experimental investigations are similar to true experiments in that both utilize treatment (TG) and comparison groups (CG) to compare outcomes between participants who receive an intervention and those who do not (Shadish et al., 2002). However, in educational settings, random assignment is often impractical. Instead, quasi-experimental designs allow researchers to study interventions in real-world conditions, where random group assignment may be impractical or introduce potential confounds (Rossi et al., 2004; Shadish et al., 2002). To mitigate these concerns, careful consideration was given to the selection of CG participants to minimize bias.

The independent variable (IV) in this study was participation or nonparticipation in the Every Day STEM for All (EDSTEMA) program. The dependent variables (DVs) were science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) in science. To measure SCK, teachers in both the TG and CG completed an SCK assessment aligned with the content covered during the EDSTEMA summer institute (SI). The SCK assessment was administered as a pretest before the SI and as a posttest after the final follow-up session. STE was assessed using the *Science Teaching Efficacy Beliefs Instrument–Form A* (STEBI-A), which was administered to both groups before and after EDSTEMA. Science SAA was measured using Oklahoma Core Curriculum Tests (OCCT) data, collected from the TG for the academic year prior to EDSTEMA (Year 0) and the year EDSTEMA was implemented (Year 1). Since no state science test is administered in fourth grade, OCCT data were only collected from fifth-grade teachers.

To evaluate the impact of EDSTEMA on SCK, STE, and science SAA, multiple statistical tests were employed to ensure robust analysis. Paired-samples *t*-tests assessed within-group changes from pretest to posttest, while independent-samples and Welch's *t*-tests compared posttest scores between the TG and CG. When the assumptions of normality or homogeneity were not met, nonparametric alternatives such as the Wilcoxon signed-rank test and Mann–Whitney *U* test were applied. In addition, analysis of covariance (ANCOVA) was conducted to control for baseline differences and further examine group effects. This multi-method approach allowed for more reliable and comprehensive analysis of the data. The following section outlines the research questions and hypotheses that guided the investigation.

Research Questions and Hypotheses

The purpose of this study was to assess the extent to which participation in the sustained, coherent, cohort-based Every Day STEM for All (EDSTEMA) professional development (PD) program influenced teacher and student outcomes. The EDSTEMA program was designed to actively engage participants in science content, with the goal of enhancing teacher science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). To guide this investigation, the following research questions (RQs) and corresponding hypotheses (Hs) were developed:

Research Questions

RQ1: Does participation in EDSTEMA significantly improve elementary teachers' science content knowledge (SCK)?

RQ2: Does participation in EDSTEMA significantly improve elementary teachers' science teaching self-efficacy (STE)?

RQ3: Does participation in EDSTEMA significantly improve student academic achievement (SAA) in science compared to the previous year?

Hypotheses

H1: Teachers who participate in EDSTEMA will demonstrate a statistically significant increase in science content knowledge (SCK) from pretest to posttest, compared to teachers who do not participate.

H2: Teachers who participate in EDSTEMA will demonstrate a statistically significant increase in science teaching self-efficacy (STE) from pretest to posttest, compared to teachers who did not participate.

H3: Students whose teachers participate in EDSTEMA will demonstrate a statistically significant increase in student academic achievement (SAA) scores in science from the year prior to EDSTEMA implementation to the year of implementation.

Population and Sample

The population for this study consisted of fourth- and fifth-grade science and math teachers from five suburban school districts within the Oklahoma City metropolitan area. While the districts are situated in relatively close proximity to one another, they exhibit considerable variation in demographic and socioeconomic characteristics (Office of Educational Quality and Accountability, 2015).

Specifically, the districts differed in size, with student enrollments ranging from 29,323 students in the smallest district to 142,265 in the largest. Socioeconomic factors also varied, with poverty rates ranging from 8–18%, and average household incomes ranging from $63,000 to $101,000. Additionally, the percentage of minority students in each district ranged from 20–58%.

The districts also varied in community characteristics. One district is located in the same city as the state’s largest university, while another is situated near a regional public university. One district is primarily considered a bedroom community, another is classified as semi-urban, and a third is considered semi-rural.

All five partner districts were mid-sized suburban districts located within commuting distance of a research university and STEM partners. This context likely facilitated recruitment, coordination of a multi-district cohort, and access to on-site and summer institute (SI) experiences. In more rural districts, longer travel distances, smaller staffs, and limited release time may complicate implementation of a similar professional development (PD) model, while large urban districts may face challenges related to scale, turnover, and competing initiatives. These contextual factors should be considered when interpreting the representativeness of the sample and the transferability of the Every Day STEM for All (EDSTEMA) design.

In sum, the participating districts provided a diverse context for examining the impact of EDSTEMA, with meaningful variation in size, socioeconomic status (SES), and community characteristics. These differences highlight the importance of understanding not only the overall population but also the specific composition of the treatment and comparison groups. The following sections describe the demographics of each group, the criteria used for study inclusion, and the ethical safeguards implemented to protect participants, including considerations of risks and benefits associated with participation.

Treatment Group

A total of 44 upper elementary math and science teachers were recruited to participate in the Every Day STEM for All (EDSTEMA) professional development (PD) program. Participation was open to all fourth- and fifth-grade science and math teachers within the partner districts. Due to high demand, applications exceeded the available spots. As a result, priority was given to early-career teachers, to teachers whose students had low Oklahoma Core Curriculum Tests (OCCT) scores, and to those who otherwise demonstrated a need to improve their science or math pedagogical skills.

Benefits and Incentives

Participants committed 86 hr of their personal time during the summer months and on weekends to attend Every Day STEM for All (EDSTEMA). In return, they received a $2,000 stipend along with books and instructional materials to support their teaching. While these incentives helped ensure strong participation and engagement, they may have also introduced selection bias, as participants were self-selected and motivated by financial and professional development benefits. This potential bias is important to consider when interpreting the study’s findings, as participants may not fully represent the broader population of elementary teachers. These characteristics of the treatment group shaped the sample composition and are described in greater detail in the following methods section.

Treatment Group Demographics

The treatment group (TG) was composed predominantly of female participants (91%). Teachers came from 26 public elementary schools in five districts, with 77% (*n* = 34) working in schools where at least half of the student body qualified for free or reduced lunch (FRL).

In terms of grade-level assignments, 48% (*n* = 21) of participants taught fourth grade, while 61% (*n* = 27) taught fifth grade; one teacher taught both grades, which explains the overlap in percentages. Regarding subject specialization, 9% (*n* = 4) of participants taught only science, 75% (*n* = 33) taught science along with one or more other subjects, and 16% (*n* = 7) taught only subjects other than science.

Participants' teaching experience spanned from one to 30 years, with an average of 11 years and a median of 8.5 years. The group included two (4.5%) first-year teachers, while 30% (*n* = 13) had between two and five years of experience. Additionally, 18% (*n* = 8) had 6–10 years of experience, 23% (*n* = 10) had 11–15 years of experience, 14% (*n* = 6) had 16–20 years of experience, 7% (*n* = 3) had 21–25 years of experience, and 5% (*n* = 2) had 26–30 years of experience. These demographics are summarized in Table 2.

Comparison Group

A total of 44 fourth- and fifth-grade science and math teachers from schools in the same five districts were recruited to serve as the comparison group (CG). Efforts were made to recruit CG members with characteristics comparable to those of the treatment group (TG). Participants in the CG were compensated $100 for each assessment they submitted, which may have introduced selection bias because participants were self-selected and potentially influenced by financial incentives.

Table 2

Participant Characteristics

Treatment Group (*n* = 44)		Comparison Group (*n* = 26)	
Gender			
Female 91% Male 9%		Female 96% Male 4%	
Percent teaching at schools where the majority of students qualify for free/reduced lunch			
77%		73%	
Grade levels taught			
4th grade 48% 5th grade 61%		4th grade 50% 5th grade 58%	
Subjects taught			
Science only 9% Science and other subjects 75% No science 16%		Science only 0% Science and other subjects 85% No science 15%	
Years of experience			
0–1	4.5%	0–1	12%
2–5	30%	2–5	27%
6–10	18%	6–10	9%
11–15	23%	11–15	27%
16–20	14%	16–20	15%
21–25	7%	21–25	4%
26–30	5%	26–30	8%

Note. Percentages may not total 100 due to rounding.

Study Inclusion Criteria

Of the 44 recruited teachers, only 26 completed both the pretest and posttest administrations of the science content understanding (SCK) assessment and/or the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) while serving as comparison group (CG) members. Consequently, data analysis was limited to these 26 participants, ensuring that only those with complete pretest and posttest data were included in this study. This approach helped maintain consistency in the dataset; however, the limited participation may have affected the study's findings, particularly regarding the representativeness of the CG.

To be eligible for the CG, teachers were required to teach at least one math or science course at the fourth- or fifth-grade level. Among participants, 50% (n = 13) taught fourth grade, 58% (n = 15) taught fifth grade, and three teachers taught both grades and were counted in both categories. In terms of subject specialization, 85% (n = 22) taught science along with one or more other subjects, while 15% (n = 4) taught only subjects other than science.

Comparison Group Demographics

Similar to the treatment group (TG), the majority of comparison group (CG) members were female (96%). Teachers were recruited from 18 public schools in the same five districts as the TG. Socioeconomic characteristics of the schools varied, with 73% (n = 19) of CG participants working in schools where at least half of the student body qualified for free or reduced lunch (FRL).

The CG exhibited a wide range of teaching experience, spanning from 1 to 30 years. The average years of experience was 10.5 years, while the median was 9.5 years. Three (12%) were first-year teachers; seven (27%) had two to five years of experience, two (9%) had six to 10 years, seven (27%) had 11 to 15 years, four (15%) had 16 to 20 years, one (4%) had 21 to 25 years, and two (8%) had 26 to 30 years. These demographics are summarized in Table 2.

Ethical Considerations

Throughout the research process, all ethical standards and guidelines outlined by the American Psychological Association (APA) and the University of Oklahoma's Institutional Review Board (IRB) were strictly followed. The research process was designed to protect participants' autonomy, privacy, and dignity. By emphasizing transparency, confidentiality, and voluntary participation, the study demonstrates a strong commitment to ethical research that respects the rights and welfare of all participants.

Researcher Positionality

In this study, I occupy multiple roles that are directly connected to the design, implementation, and evaluation of the Every Day STEM for All (EDSTEMA) program. I was responsible for securing the grant that funded EDSTEMA, leading the development of the program, administering it with participating districts and teachers, and conducting the analyses reported in this dissertation. As a result, I am not a neutral, external observer; rather, I am deeply invested in EDSTEMA's success.

This insider position affords important strengths. My involvement in conceiving and developing EDSTEMA provided a nuanced understanding of the program's goals, structures, and underlying theory of change, which informed the formulation of the research questions, the selection of instruments, and the interpretation of findings. My ongoing relationships with participating teachers and district partners also facilitated access to the field and helped build the trust necessary for collecting sensitive data about beliefs, practices, and student outcomes.

At the same time, my multiple roles introduce potential sources of bias. I have a professional and personal interest in EDSTEMA being perceived as effective, which could influence how I frame the study, interpret ambiguous results, or foreground certain findings over others. The fact that I administered the program may also have shaped how teachers responded to surveys or engaged in program activities, given possible power dynamics and perceptions of evaluation.

To address these concerns, I approached the study with an explicit commitment to reflexivity and transparency. I relied on established, validated instruments where possible, used predefined analytic procedures aligned with the research questions, and documented analytic decisions throughout the process. Wherever feasible, I triangulated across multiple data sources (e.g., teacher self-report, content assessments, and student achievement data) to corroborate findings rather than relying on a single indicator of impact. I also engaged in ongoing reflection about how my commitments to EDSTEMA might shape my interpretations and sought feedback from colleagues on preliminary analyses and conclusions. These strategies do not eliminate the influence of my positionality, but they are intended to make it visible and to strengthen the trustworthiness of the study.

Institutional Review Board (IRB) Approval

Prior to data collection, Institutional Review Board (IRB) approval was obtained from the University of Oklahoma to ensure that the study adhered to ethical standards for research involving human participants. As part of this process, the IRB reviewed the study's overall design, proposed procedures, and informed consent documents to confirm compliance with all established research guidelines. This approval process safeguarded participants by ensuring that issues of autonomy, confidentiality, and voluntary participation were appropriately addressed. The IRB approval number for this study is 5393, reflecting formal authorization to proceed with data collection under these protections.

Risks and Benefits of Participation

This study posed minimal risks to participants. The primary activities involved completing surveys and participating in professional development (PD) sessions. While these activities are generally considered low-risk, participants were informed that the study required a substantial time commitment. This time commitment was identified as the primary risk, as it required teachers to dedicate personal time during the summer and school year to fully engage in the Every Day STEM for All (EDSTEMA) program. Participants were assured that their involvement was voluntary and that they could withdraw at any point without penalty.

The study provided both educational and tangible benefits. The most significant benefits were educational, as participants had the opportunity to deepen their science content understanding (SCK) and strengthen their science teaching self-efficacy (STE). By engaging in EDSTEMA, teachers were able to enhance their instructional practices and build confidence in teaching science, outcomes that not only supported their professional growth but also had the potential to improve student learning.

In addition to these educational benefits, tangible incentives were also provided. Teachers in the treatment group (TG) received a $2,000 stipend, along with books and instructional materials for use in their classrooms. Teachers in the comparison group (CG) were compensated $100 for each assessment they completed. These incentives helped encourage participation but may also have introduced some degree of self-selection bias, as individuals motivated by financial benefits may have been more likely to participate.

Overall, the risks associated with this study were minimal, while the potential benefits—both educational and material—were substantial. The combination of professional growth opportunities, classroom resources, and financial compensation made participation in EDSTEMA a valuable experience for many teachers. Importantly, all risks and benefits were clearly communicated during the informed consent process, and the study was reviewed and approved by the University of Oklahoma IRB to ensure compliance with ethical research standards as outlined by the American Psychological Association (APA).

Informed Consent

All participants received an informed consent form outlining the study's purpose, procedures, potential risks and benefits, and the voluntary nature of participation. Participants were informed that they could withdraw from the study at any point without consequences. The informed consent document was written in clear, accessible language, and participants were given opportunities to ask questions before signing. By signing, participants indicated their voluntary consent to take part in the study.

If a participant chose to withdraw, their data were discarded and excluded from all analyses. This procedure ensured that participants' autonomy was respected and that only data from those who fully consented were included in the final results. The informed consent process adhered to ethical standards established by the University of Oklahoma IRB and the American Psychological Association (APA), underscoring the study's commitment to transparency, participant rights, and ethical research practices.

Confidentiality and Data Protection

To protect participants' privacy and ensure confidentiality, all personally identifying information was removed from the data before analysis. Electronic research files were stored in password-protected formats and accessible only to the researchers directly involved in the study. Paper-based documents, such as consent forms and assessment records, were kept in a locked, secure location to further safeguard participant information. These measures ensured that participants' identities remained protected throughout the study and reinforced the researchers' commitment to maintaining the highest ethical standards in human subjects research.

Population and Sample Summary

In sum, this study involved a sample of fourth- and fifth-grade math and science teachers drawn from five diverse suburban districts in the Oklahoma City metropolitan area. Although both treatment group (TG) and comparison group (CG) were recruited to ensure comparability, incentives and self-selection may have limited representativeness. Nonetheless, careful adherence to ethical standards—including informed consent, confidentiality, and Institutional Review Board (IRB) approval—helped ensure that participants' rights were safeguarded and that data collection adhered to rigorous research protocols. Having outlined the study population and ethical considerations, the next section describes the Every Day STEM for All (EDSTEMA) program, detailing its structure, goals, and design features as the central intervention under investigation.

EDSTEMA Program

The name Every Day STEM for All (EDSTEMA) reflects the belief that every student—regardless of socioeconomic background, race, ethnicity, gender, disability, or language proficiency—deserves access to a high-quality science, technology, engineering, and mathematics (STEM) education that includes science instruction, every day. EDSTEMA was a year-long professional development (PD) initiative for fourth- and fifth-grade science and math teachers. The program began with a two-week summer institute (SI) and included three follow-up sessions throughout the academic year. Its primary goal was to strengthen teachers' science content understanding (SCK), instructional skills, and confidence. Following the SI, participants were expected to design and implement lessons that integrated the content and strategies introduced during the training.

EDSTEMA was deliberately structured as a comprehensive, year-long PD initiative aligned with research-based features of effective PD (Desimone, 2009): sustained duration, collective participation, content focus, active learning, and coherence. Each of these elements was integrated into the EDSTEMA program to maximize its effectiveness in improving teachers' SCK, science teaching self-efficacy (STE), and pedagogical content knowledge (PCK) and instructional practices, while ultimately improving student academic achievement (SAA) in science.

In addition to aligning with general features of effective PD, EDSTEMA's structure was intentionally designed to support efficacy formation in Bandura's (1977) sense by providing teachers with repeated mastery experiences in science content, opportunities to observe expert and peer models, and ongoing feedback and encouragement within a supportive professional community.

To ensure that the program addressed the most pressing challenges in partner districts, its design was informed by a systematic needs assessment (NA), followed by intentional structuring of the summer institute SI, follow-up sessions, and lesson development. The next sections describe the needs assessment that shaped the program and outline the specific design elements that guided EDSTEMA's implementation.

Needs Assessment

The Every Day STEM for All (EDSTEMA) program was created in response to the need for improving elementary science education in the partner districts. To ensure the program design was grounded in local context, facilitators first conducted a needs assessment (NA) to identify the specific professional development priorities of upper-elementary teachers. The results of this NA directly informed the structure of the program, ensuring that it addressed the most critical challenges.

Findings from the NA indicated that partner school districts faced persistent challenges in delivering high-quality science education to all elementary students. Multiple data sources informed this conclusion, including interviews with middle school science teachers, analysis of fifth-grade science Oklahoma Core Curriculum Tests (OCCT) scores, results from an science teaching self-efficacy (STE) inventory, a survey on amount of time spent on science instruction and methods of science instruction at the upper elementary level, and an science content understanding (SCK) assessment. Together, these data highlighted key areas of concern that shaped the program's design. The following sections outline the major NA findings and illustrate how they guided the development of EDSTEMA's professional development model.

Middle School Teacher Interviews

To assess middle school teachers' perceptions of student readiness for middle school science, Every Day STEM for All (EDSTEMA) facilitators administered a survey to middle school teachers in the partner districts (see Appendix A). Participation was voluntary, and all middle school teachers were invited to respond. Middle school was selected as the focus of this inquiry because in many schools within the partner districts, it marked the beginning of departmentalized instruction, where students first experienced dedicated science instructional time. As a result, EDSTEMA facilitators sought to determine whether middle school science teachers believed incoming students were adequately prepared for grade-level science instruction.

Results from the middle school teacher survey revealed that incoming students were viewed as poorly prepared for grade-level science, with all 18 respondents expressing dissatisfaction and providing an average rating of just 1.83 on a 5-point scale. This finding indicated that elementary students were entering middle school without mastering essential science knowledge and skills, leaving teachers to remediate foundational content rather than advance learning. These results highlighted a critical gap in elementary science instruction and underscored the urgent need for interventions like EDSTEMA that aimed to strengthen teachers' capacity to build stronger science foundations at earlier grades.

In addition to the quantitative ratings, teachers elaborated on their concerns in open-ended responses. Teachers were asked to describe their perceptions about how well-prepared incoming students were for the middle school science curriculum. One teacher commented:

> I receive students fresh from elementary school with little to no science experience or knowledge. Many may have done some science but have limited knowledge in their understanding, with many having no experience, excluding reading about science. I have to introduce the basic concepts of science; essentially, I am building the foundation, not building on a foundation in science understanding.

Another respondent stated:

> [Entering middle schoolers'] background is unsatisfactory. At the seventh-grade level, we spend much of our time covering material that should've been mastered at the elementary level rather than extending the [science] skills and standards at an appropriate higher level.

Together, these quantitative and qualitative findings underscored the significant need for improved science instruction at the elementary level, reinforcing the rationale for EDSTEMA's focus on strengthening science teaching in the early grades.

Fifth-Grade Science Oklahoma Core Curriculum Test Scores

The fifth-grade science Oklahoma Core Curriculum Tests (OCCT) scores (summarized in Table 3 and presented in full in Appendix B) highlighted the urgent need to strengthen elementary science instruction. In each partner district, fewer than 70% of students demonstrated proficiency. This was concerning, as research has consistently shown that early proficiency in science is positively correlated with achievement in middle and high school, and success in these later grades predicts postsecondary readiness (ACT, 2008; Harris Interactive, 2011; Traphagen, 2011).

These results indicated that a substantial proportion of students in the partner districts were entering secondary school without a strong foundation in science. Because elementary teachers play a pivotal role in developing this foundation, improving science instruction in the early grades was essential. Strengthening elementary science education was therefore a critical step toward promoting scientific literacy, raising achievement in subsequent grades, and expanding students' long-term academic and career opportunities.

Science Teaching Self-Efficacy Needs Assessment

Bandura's (1977) self-efficacy theory posits that an individual's beliefs about their abilities in a given area influence their performance in that domain. Research shows that teachers with high science teaching self-efficacy (STE) demonstrate stronger instructional performance and science student academic achievement (SAA) compared to those with low STE. Conversely, teachers with low STE are more likely to avoid teaching science and provide lower-quality instruction (Ashton, 1984; Bandura, 1982; Riggs, 1988; Sandholtz & Ringstaff, 2014).

Table 3

Percent Scoring Proficient or Above on Fifth-Grade Science OCCT by District and Demographic

	District 1	District 2	District 3	District 4	District 5
Race and Ethnicity					
Overall	60	68	64	57	49
Hispanic/Latino	59	60	62	39	55
American Indian/Alaskan Native	55	89	65	42	71
Asian	n/a	92	82	87	79
Black/African American	53	45	61	39	38
Pacific Islander	n/a	75	100	0	33
White/Caucasian	67	78	74	69	65
Two or more races	72	74	72	71	53
Gender					
Female	67	72	71	63	53
Male	65	75	72	66	60
Socioeconomics					
Economically disadvantaged	55	54	59	50	49
Non-economically disadvantaged	72	80	81	75	71
English Language Proficiency					
English Language Learners (ELL)	0	29	36	23	19
ELL first year proficient	n/a	14	70	42	46
ELL second year proficient	n/a	63	49	54	58
Non-English language learners (non-ELL)	60	75	66	59	52
Students with Individualized Education Programs (IEPs)					
IEP with accommodations	26	29	25	24	23
IEP without accommodations	70	51	53	44	47
Individualized Education Program (IEP)	33	38	34	30	29
Regular education	66	74	72	64	56

Note: Includes only students enrolled for the full academic year. OCCT = Oklahoma Core Curriculum Test.

Science Teaching Self-Efficacy Needs Assessment

Bandura's (1977) self-efficacy theory posits that an individual's beliefs about their abilities in a given area influence their performance in that domain. Research shows that teachers with high science teaching self-efficacy (STE) demonstrate stronger instructional performance and science student academic achievement (SAA) compared to those with low STE. Conversely, teachers with low STE are more likely to avoid teaching science and provide lower-quality instruction (Ashton, 1984; Bandura, 1982; Riggs, 1988; Sandholtz & Ringstaff, 2014).

To assess the level of STE among elementary teachers in the partner districts, the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) was administered to upper elementary science teachers (*n* = 19). All fourth- and fifth-grade teachers in the partner districts were invited to participate, with participation being voluntary. The STEBI-A consists of 25 Likert-scaled items, each rated on a five-point scale: 1– *strongly disagree*, 2– dis*agree*, 3– *uncertain*, 4– *agree*, and 5– *strongly agree*. Items 3, 6, 8, 10, 13, 17, 19–22, 24, 25 are negatively worded and were reverse-coded to maintain consistency, so that *strongly agree* received a score of 1 and *strongly disagree* was assigned a score of 5 (Riggs, 1988). The mean response values and range of responses were recorded for each (see Table 4).

Specifically, response averages for items 3, 5, 10, 12, 15, 17, 18, 19, 25 indicate that some fourth- and fifth-grade science teachers had lower STE. Responses to items 3, 5, 12, 18, and 19 suggest that this lower STE may stem—at least partly—from a lack of science content understanding (SCK) and science pedagogical skills. Based on Bandura's (1977) theory that teacher self-efficacy influences student achievement, the STEBI-A results suggest that strengthening teachers' SCK and pedagogical skills may be an effective strategy for improving elementary students' science achievement.

Table 4

STEBI-A: Instrument and Needs Assessment Results

	Statement	Average score	Range of responses
1.	When a student does better than usual in science, it is often because the teacher exerted a little extra effort.	4.06	2–5
2.	I am continually finding better ways to teach science.	4.25	2–5
3.	Even when I try very hard, I don't teach science as well as I do most subjects.	3.94	1–5
4.	When the science grades of students improve, it is most often due to their teacher having found a more effective teaching approach.	4	3–5
5.	I know the steps necessary to teach science concepts effectively.	3.06	1–5
6.	I am not very effective in monitoring science experiments.	4	2–5
7.	If students are underachieving in science, it is most likely due to ineffective science teaching.	3.5	1–5
8.	I generally teach science ineffectively.	2	1–3
9.	The inadequacy of a student's science background can be overcome by good teaching.	4	3–5
10.	The low science achievement of some students cannot generally be blamed on their teachers.	4.44	2–5
11.	When a low achieving child progresses in science, it is usually due to the extra attention given by the teacher.	3.75	2–5
12.	I understand science concepts well enough to be effective in teaching elementary science.	3.56	1–5
13.	Increased effort in science teaching produces little change in some students' science achievement.	2.06	1–4
14.	The teacher is generally responsible for the achievement of students in science.	3.62	2–5
15.	Students' achievement in science is directly related to their teacher's effectiveness in science teaching.	3.56	2–5
16.	If parents comment that their child is showing more interest in science at school, it is probably due to the performance of the child's teacher.	4	3–5
17.	I find it difficult to explain to students why science experiments work.	2.44	1–5
18.	I am typically able to answer students' science questions.	3.56	2–5
19.	I wonder if I have the necessary skills to teach science.	2.87	1–5
20.	Effectiveness in science teaching has little influence on the achievement of students with low motivation.	2.87	1–5
21.	Given a choice, I would not invite the principal to evaluate my science teaching.	2.31	1–5
22.	When a student has difficulty understanding a science concept, I am usually at a loss as to how to help the student understand it better.	2.37	1–5
23.	When teaching science, I usually welcome student questions.	4.66	3–5
24.	I don't know what to do to turn students on to science.	2.19	1–4
25.	Even teachers with good science teaching abilities cannot help some kids learn science.	3.06	1–5

Note: Response scale: 1 = *strongly disagree*, 2 = *disagree*, 3 = *uncertain*, 4 = *agree*, 5 = *strongly agree*. STEBI-A = Science Teaching Efficacy Beliefs Instrument–Form A.

Time Spent on Science Instruction Survey

Seventeen fourth- and fifth-grade teachers from partner districts completed a voluntary survey in which they reported the number of minutes per week spent teaching science. Responses varied widely, ranging from 0 to 225 min per week, with an average of 107 min. This equates to approximately two 55-min periods per week, or about 21 min per day.

In contrast, elementary teachers were required to allocate 90 min per day to English language arts (ELA) instruction, as mandated by the Reading Sufficiency Act (RSA, 1997). These survey findings were consistent with informal interviews, in which teachers reported that most of the school day was dedicated to ELA and math. After accounting for lunch, recess, and *specials* (e.g., art, physical education, music), science had to compete for the limited remaining instructional time with ELA and math remediation, as well as social studies.

Together, these results align with national trends showing that science instruction at the elementary level receives significantly less time than math and ELA and has steadily declined over the past two decades (Blank, 2013; Trygstad et al., 2013). These conditions underscore the importance of initiatives like Every Day STEM for All (EDSTEMA), which were designed to expand opportunities for elementary students to engage in meaningful science learning.

Methods of Science Instruction Survey

A strong case has been made for the use of inquiry-based teaching methods in science instruction, as opposed to traditional approaches. Traditional teaching methods are typically teacher-centered, with the teacher determining the problem to be investigated, the methods of investigation, and how information is analyzed (Domin, 1999).

In contrast, inquiry-based approaches are more student-centered and involve engaging learners in scientific questions, allowing them to design and conduct investigations, formulate explanations, and communicate and justify their findings (National Research Council [NRC], 2000). Research indicates that inquiry-based instruction is more effective than traditional methods in promoting student understanding of science content and practices (Minner et al., 2010).

To better understand how these patterns manifested locally, the Every Day STEM for All (EDSTEMA) needs assessment survey asked teachers to describe their typical methods of science instruction. Of the 17 respondents, 15 described their typical methods of science instruction, while two left the item blank. Ten respondents indicated that they engaged students in inquiry-based science, whereas only two provided explanations of what they considered to be *inquiry*.

Both practical experience and research suggest that teachers often believe they are implementing inquiry-based strategies when, in reality, they are not (Capps & Crawford, 2013). Studies have shown that traditional teaching methods remain the most prevalent methods of science instruction. One comprehensive study found that lecture, textbook reading, and answering textbook questions are more commonly used in elementary than hands-on laboratory activities, which may or may not incorporate inquiry (Trygstad et al., 2013).

Findings from the EDSTEMA NA survey further supported this trend. Eleven of the 15 respondents reported that reading a textbook, taking notes, and answering textbook questions were primary methods of instruction in their science classes. Teachers in the partner districts reported an underutilization of inquiry-based methods and an overreliance on traditional approaches.

These findings were concerning, as inquiry-based instruction has been shown to be more effective for science learning. This reliance on less effective instructional methods highlighted the urgent need for professional development (PD) programs that aimed to strengthen elementary teachers' capacity to implement effective, inquiry-based science pedagogy.

Science Content Knowledge Needs Assessment

In order to determine participants' level of science content understanding (SCK), selected questions from the *Misconceptions-Oriented Standards-Based Assessment Resources for Teachers* (MOSART) were administered. To guide this selection, partner districts provided fifth-grade science Oklahoma Core Curriculum Tests (OCCT) scores, which were analyzed to identify areas of low student performance. These content areas were prioritized, as they were likely to be focal points of instruction during the Every Day STEM for All (EDSTEMA) summer institute. In addition, because the state was set to adopt new science standards the following year, the team reviewed upcoming fourth- and fifth-grade standards to identify topics that would represent new instructional content. MOSART items addressing these areas were added to the pool of potential questions.

Because elementary science teachers are responsible for teaching content standards across all science domains, the resulting assessment included a balanced combination of items from space, life, physical, and Earth science. To minimize the risk of participant fatigue and ensure completion, only selected questions were used instead of the full 85-question instrument. The final exam consisted of 34 representative questions aligned with the National Science Education Standards (NSES). All fourth- and fifth-grade teachers in partner districts were invited to participate, with participation being voluntary.

Following administration, the average percent correct per question and per domain was calculated (see Appendix C). The overall average score for all four sections of the abridged MOSART was 62.9%. Among the domains, Earth science had the highest average score

(72.6%), followed by life science (63.3%), physical science (45%), and space science (43.3%). These results suggest that the participants lacked SCK up to three or four grade levels beyond that which they were responsible for teaching.

Taken together with other data sources, the MOSART results reinforced a consistent pattern: elementary teachers in the partner districts struggled with science content knowledge across multiple domains. These findings, combined with survey data, teacher interviews, and assessment results, painted a clear picture of the challenges facing elementary science instruction.

The collective evidence from multiple sources—including middle school teacher interviews, fifth-grade science OCCT scores, survey results on instructional time and teaching methods, Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) responses, and MOSART scores—highlighted the urgent need for high-quality professional development in elementary science content and pedagogy. Middle school teachers reported that students entered unprepared for grade-level science, a concern echoed by below-average OCCT scores. Limited instructional time and reliance on traditional methods likely contributed to these outcomes.

Consistent with Bandura’s (1977) theory, results from the STEBI-A suggested that gaps in content knowledge (CK) and pedagogical skills undermined teachers’ self-efficacy (TSE), while MOSART scores confirmed weaknesses in CK. Together, these findings informed the design of the EDSTEMA program, ensuring that it directly addressed the science content and pedagogical needs of teachers in the partner districts (Appendix D).

EDSTEMA Design

The Every Day STEM for All (EDSTEMA) program arose out of the need to address these documented gaps in elementary science preparation. The clear evidence that students were entering middle school without adequate science foundations underscored the importance of strengthening teacher science content understanding (SCK), pedagogical content knowledge (PCK), and science teaching self-efficacy (STE) at the elementary level.

To meet this challenge, EDSTEMA was designed as a year-long professional development (PD) initiative that directly responded to identified needs, aligning its structure with research-based features of effective PD. The following section describes how the program was intentionally designed to address these challenges and support lasting improvements in science teaching and learning.

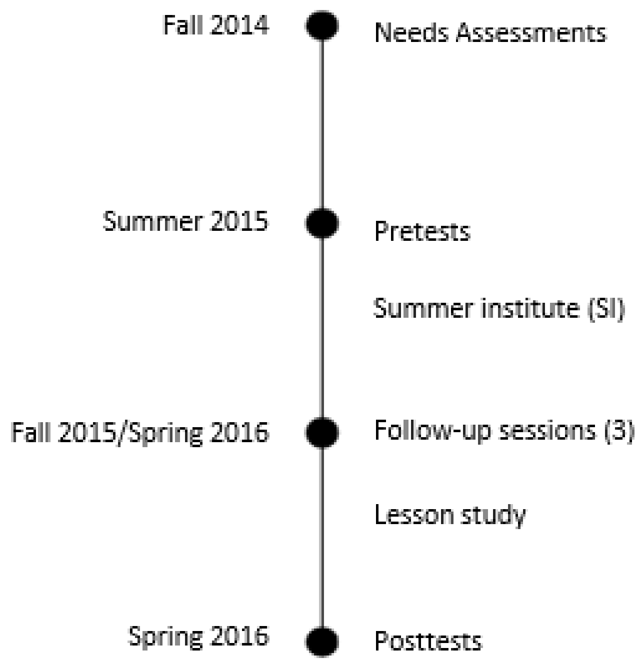
EDSTEMA was designed for elementary teachers in collaboration with STEM faculty from five suburban metro school districts, one large public state university, one small private Catholic university, and one university research center. The program provided 86 hr of PD to 44 fourth- and fifth-grade science and/or math teachers from partner districts. The program was structured into three phases: (a) an intensive two-week summer institute (SI), (b) three follow-up day-long PD workshops held throughout the school year, and (c) lesson development work, as shown in the timeline in Figure 3. This year-long format ensured sustained support and multiple opportunities for teachers to deepen their SCK and strengthen instructional practices.

Summer Institute

During the summer institute (SI), STEM faculty—who possessed strong science content understanding (SCK), expertise in inquiry-based teaching strategies, and extensive experience in facilitating teacher professional development (PD)—engaged teachers in inquiry-based science lessons. These lessons were designed to increase teachers' SCK while simultaneously modeling effective instructional strategies.

Figure 3

Timeline of Every Day STEM for All (EDSTEMA) activities



Note. The timeline illustrates the sequence of key activities in the EDSTEMA professional development program, including the initial Summer Institute, follow-up sessions during the academic year, and subsequent data collection phases.

Because the needs assessment (NA) results indicated that fourth- and fifth-grade teachers in the participating districts performed poorly on SCK assessments designed for students in grades five through eight across the four science domains (space, life, physical, and Earth), the SI was structured to address these deficiencies. Specifically, participants engaged in activities aimed at strengthening their SCK in each domain beyond the scope of what they were expected to teach.

To guide content selection, Every Day STEM for All (EDSTEMA) facilitators identified key topics from a *Content Priority List* (Appendix E). This list included Oklahoma Academic Standards (OAS) topics in which:

- fifth-grade students demonstrated low proficiency on the Oklahoma Core Curriculum Test (OCCT),
- teachers exhibited low SCK on the Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) assessments, and/or
- the content was newly introduced under the revised OAS and had not been required under the previous Priority Academic Student Skills (PASS) standards, resulting in limited teacher familiarity.

These identified topics formed the basis for the inquiry-based lessons presented during the SI and follow-up PD sessions. The EDSTEMA facilitators implemented an inquiry-based instructional approach to teach these concepts, as research indicates that the most effective PD engages teachers in active learning experiences, where subject matter knowledge is closely integrated with its practical application in the classroom (Shulman, 1992).

Follow-Up Sessions

Three full-day follow-up sessions were held on Saturdays during the academic year. Each session focused on two major topics designed to extend teachers' learning from the Summer Institute and support classroom implementation. Session topics included data-driven instruction, inquiry and the 5E model, science–math integration, evaluating STEM lessons, literacy in STEM, and sustainability planning.

In the data-driven instruction session, Every Day STEM for All (EDSTEMA) facilitators guided teachers in analyzing Oklahoma Core Curriculum Test (OCCT) scores and formative assessments to align their curricula with grade-level standards. The inquiry session introduced the 5E instructional model and provided time for teachers to design their own 5E lessons. The science–math integration and EQuIP (Educators Evaluating the Quality of Instructional

Products) rubric sessions helped participants evaluate and adapt lessons for content alignment and instructional quality. A literacy in STEM session modeled strategies for scaffolding student writing using the Claims–Evidence–Reasoning (CER) framework. Finally, the sustainability session supported teachers in developing action plans to maintain and disseminate the Every Day STEM for All approach within their schools. Key session objectives are summarized in Table 5.

Lesson Development

In addition to the summer institute (SI) and follow-up sessions, lesson development was an essential component of the Every Day STEM for All (EDSTEMA) model. The lesson development component required teachers—working individually, in pairs, or in groups—to modify an existing lesson or develop a new lesson using the pedagogical and assessment strategies introduced during the EDSTEMA experience. Participants were expected to implement the lesson during the school year and report student pre- and post-assessment data to EDSTEMA facilitators. This process allowed teachers to apply the science content understanding (SCK) and instructional strategies learned during the SI and follow-up sessions.

To support lesson development, participants were encouraged to collaborate with EDSTEMA facilitators and receive feedback on key aspects of their lessons, including the integration of science and math, increasing instructional rigor, and incorporating strategies for English learners (ELLs) and students with Individualized Education Programs (IEPs). EDSTEMA facilitators observed volunteer participants as they delivered their lessons and conducted post-lesson debriefing sessions to provide constructive feedback. At the conclusion of the final follow-up session, participants completed post-assessments of both SCK and science teaching self-efficacy (STE) to evaluate retention and long-term changes.

Table 5

Description of Follow-Up Session Activities

Session	Description
Session 1a: Data-Driven Instruction: Interpreting Student Performance and Aligning Curriculum	EDSTEMA facilitators guided teachers in analyzing OCCT scores and formative assessment results and provided strategies for interpreting the standards for each grade level. Teachers used this information to align their curricula with the required standards.
Session 1b: Inquiry and the 5Es	Facilitators introduced the characteristics of effective inquiry-based lessons and presented the 5E instructional model. Teachers had the opportunity to develop their own 5E lessons by the end of the session.
Session 2a: Strategies for Science–Math Integration	Participants explored a framework and rubric for evaluating the extent to which lessons effectively integrate both science and math.
Session 2b: A Guide for Selecting and Modifying STEM Lessons	Teachers were introduced to the EQuIP rubric for science, which provides guidance for selecting and refining high-quality STEM lessons.
Session 3a: Literacy in STEM	Facilitators presented strategies for incorporating reading and writing into science instruction. In particular, they demonstrated strategies for scaffolding students' use of the CER model to support scientific writing and argumentation.
Session 3b: Making it Last: Planning for Sustainability	Participants developed action plans for sustaining the EDSTEMA approach in their own classrooms. Additionally, they developed strategies for sharing their newly acquired knowledge with non-participating teachers.

Note: EDSTEMA = Every Day STEM for All; OCCT = Oklahoma Core Curriculum Test; EQuIP = Educators Evaluating the Quality of Instructional Products; STEM = science, technology, engineering, math; CER = claims–evidence–reasoning

EDSTEMA Program Summary

In sum, the Every Day STEM for All (EDSTEMA) program was intentionally designed to address the pressing needs identified through the comprehensive needs assessment (NA) while aligning with research-based principles of effective professional development (PD). Its structure provided teachers with sustained opportunities to strengthen their science content understanding (SCK), enhance pedagogical content knowledge (PCK), and build science teaching self-efficacy

(STE). Table 6 summarizes the three phases of the EDSTEMA program and their corresponding goals and activities. By combining inquiry-based pedagogy, content focus, and collaborative participation, EDSTEMA created a professional learning (PL) environment aimed at producing lasting improvements in both teacher practice and student academic achievement (SAA). The following section describes the data collection procedures used to evaluate the program's impact on teacher outcomes and student learning.

Data Collection

All participants in Every Day STEM for All (EDSTEMA)—including those in both the treatment group (TG) and comparison group (CG)—provided informed consent prior to the collection and use of data for research purposes. Participants were thoroughly informed about the study's purpose, procedures, potential risks, and benefits, ensuring that they understood their rights as research subjects. They were informed that their participation was voluntary and that they could withdraw from the study at any point without penalty. Additionally, all collected data were maintained with strict confidentiality and used solely for the purposes of this study, in full compliance with ethical research guidelines and Institutional Review Board (IRB) requirements.

This section details the instruments used to measure the study's key variables, including science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). For each instrument, information is provided on its validity, reliability, administration procedures, scoring methods, and known limitations. Together, these descriptions establish the rigor of the data collection process and clarify how each measure was used to evaluate the impact of EDSTEMA.

Table 6

Summary of Every Day STEM for All (EDSTEMA) Phases

Phase	Description of Activities and Objectives
Summer Institute (SI)	The SI served as the foundational phase of the EDSTEMA program, providing an intensive, multi-day experience designed to strengthen teachers' science content knowledge and pedagogical content knowledge. Participants engaged in hands-on, inquiry-based investigations aligned with state standards, explored interdisciplinary connections among science, mathematics, and literacy, and examined effective instructional models such as the 5E framework. Collaborative lesson design and reflective discussions allowed teachers to connect new learning to classroom practice and to establish a shared foundation for the subsequent phases of professional development.
Follow-Up Sessions	Three full-day follow-up sessions were held on Saturdays throughout the academic year to extend and reinforce learning from the Summer Institute. Each session included two focused modules that deepened participants' understanding of data-driven instruction, inquiry-based teaching, interdisciplinary lesson design, and literacy integration in STEM. Facilitators guided teachers in analyzing student performance data, developing lessons that integrated science and mathematics, evaluating instructional quality using the EQuIP rubric, and scaffolding student writing through the Claims–Evidence–Reasoning model. The final session emphasized sustainability planning, during which teachers created strategies for maintaining and disseminating the Every Day STEM for All approach within their schools.
Lesson Study Cycle	The lesson study phase provided teachers with a structured opportunity to apply their learning in authentic classroom contexts through collaborative planning, teaching, observation, and reflection. Working in teams, participants designed and implemented research lessons, collected and analyzed student evidence, and engaged in reflective dialogue to identify strengths and areas for improvement. This process emphasized active learning, peer collaboration, and continuous refinement of instructional practices, promoting sustained growth and alignment between professional learning and classroom implementation.

Note. Table summarizes the major phases and activities of the EDSTEMA professional development program. SI = Summer Institute. STEM = science, technology, engineering, mathematics. EQuIP = Educators Evaluating the Quality of Instructional Products.

This section details the instruments used to measure the study's key variables, including science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). For each instrument, information is provided on its validity, reliability, administration procedures, scoring methods, and known limitations. Together, these descriptions establish the rigor of the data collection process and clarify how each measure was used to evaluate the impact of EDSTEMA.

Instruments

Three instruments were used to measure the dependent variables (DVs) of science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). Each instrument was carefully selected to ensure alignment with the goals of Every Day STEM for All (EDSTEMA) and to capture different but complementary aspects of teacher and student outcomes. The following section describes each instrument in detail, including its structure, purpose, and relevance to this study. In addition, the rationale for selecting these particular instruments and the ways in which they were administered and analyzed are explained to provide transparency and support the validity of the research design.

MOSART for Science Content Knowledge Assessment

To assess participants' science content understanding (SCK), selected questions from Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) assessments in space science, life science, physical science, and Earth science were administered (see Appendix F for questions). Developed by researchers in the Harvard–Smithsonian Center for Astrophysics Science Education Department, MOSART assessments are designed to evaluate common science misconceptions aligned with the National Science Education Standards (NSES) (The President and Fellows of Harvard College, 2007, 2011; NRC, 1996). These assessments were selected because they are reliable measures of student understanding, especially in identifying misconceptions, and have been validated in previous studies (Sadler et al., 2010).

Reliability and Validity. Although MOSART assessments have not been explicitly validated for use with adults, they were selected for this study due to the lack of a validated instrument specifically designed to assess science content understanding (SCK) or misconceptions specifically in elementary educators.

The MOSART instruments were developed through a rigorous, research-based process that included expert review of item content, alignment with National Science Education Standards (NSES), pilot testing, and field testing with thousands of students (Sadler et al., 2010). Reported Kuder-Richardson coefficients of reliability (*KR*–20) ranged from .73 to .85 across grade-levels, demonstrating acceptable to high reliability. Validity was established through expert confirmation of content accuracy, alignment with standards, and consistency of patterns of performance across age groups, all of which support the assessments' validity. These characteristics, along with the instrument's focus on diagnosing misconceptions and its close alignment with Every Day STEM for All (EDSTEMA) content, made MOSART a suitable choice for this study. However, because the assessments were originally designed for K–12 students, their use with adult educators may introduce limitations related to item difficulty and contextual relevance.

Development. MOSART assessments are available at three grade-level bands: K–4, 5–8, and 9–12. They cover four science domains: space, life, physical, and Earth sciences. For this study, the 5–8 and 9–12 grade-level assessments were administered to elementary teachers participating in Every Day STEM for All (EDSTEMA), as teachers are expected to have science content understanding (SCK) beyond the grade level which they teach. Because elementary science teachers are responsible for instructing across all four science domains, selected questions from each domain-specific assessment were administered.

To minimize participant fatigue and ensure test completion, only a subset of questions was used rather than the full 85-question assessment. One question per relevant National Science Education Standards (NSES) standard addressed in EDSTEMA activities was selected, resulting in a 34-question assessment covering space, life, physical, and Earth sciences. This approach ensured that the assessment was both aligned with content and time-efficient, balancing validity with practical constraints.

Administration. Treatment group (TG) participants completed the science content understanding (SCK) assessment pretest on the first day of Every Day STEM for All (EDSTEMA), before engaging in any professional development (PD) activities. The posttest was administered at the conclusion of the final follow-up session, after all PD activities had been completed. Comparison group (CG) participants received a digital link to the pretest at approximately the same time as the assessment was administered to TG participants and were given two weeks to complete and submit their responses. The posttest was administered in the same manner, following the completion of EDSTEMA.

Scoring. Each correct response was scored as 1, while incorrect responses were scored as 0. Respondents' total number correct was summed to generate an overall science content understanding (SCK) score. To assess knowledge gains, the change in average score from pretest to posttest was calculated for each respondent. If a respondent failed to complete both the pretest and posttest, pairwise deletion was applied to remove the missing data from the analysis. Pairwise deletion was preferred over listwise deletion in this case, as the absence of these scores did not affect other variables in the study (Shadish et al., 2002). Listwise deletion was not used because it would have reduced the sample size significantly, potentially introducing bias and decreasing statistical power. However, pairwise deletion assumes that missing data are missing at random (MAR), which may introduce some limitations if differences exist between respondents who completed both tests and those who did not.

Delimitations. Several decisions regarding the use of the MOSART assessments were intentional delimitations of this study. First, only items from the 5–8 and 9–12 grade-level

MOSART assessments were selected, as these were judged to be most appropriate for evaluating elementary teachers' science content understanding (SCK) beyond the grade level they teach. The instrument was further delimited to the four core science domains—space, life, physical, and Earth sciences—to align with the content emphasized in Every Day STEM for All (EDSTEMA). Additionally, only a selection of 34 items was used to balance content coverage with constraints such as testing time and respondent fatigue. These choices were made to ensure the assessment was both aligned with program content and feasible within the professional development (PD) schedule. The decision to administer the assessment only to teachers participating in EDSTEMA and comparison group (CG) teachers from the same partner districts also allows for the generalizability of findings to similar contexts and populations.

Limitations. One limitation of this study is that MOSART assessments have not been explicitly validated for adult learners, particularly elementary educators. While these assessments have demonstrated strong reliability and validity in K–12 student populations, their use with adult teachers may present challenges, as the nature of misconceptions in adults may differ from those of students. The application of MOSART assessments to adult educators may therefore limit the accuracy of assessing teachers' science content understanding (SCK), as the instrument was originally designed for children.

MOSART Summary. In summary, the MOSART assessment served as a practical and reliable measure of teachers' science content understanding (SCK) across the four science domains, even though it was originally developed for K–12 students. By selecting a subset of items aligned with Every Day STEM for All (EDSTEMA) content, the instrument provided focused coverage of relevant standards while minimizing participant fatigue. Administering the assessment both before and after the program allowed the study to capture growth in teachers' conceptual understanding, offering insight into the effectiveness of the professional development (PD) model. Building on this measure of content knowledge, the next section turns to the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A), which assessed teachers' confidence in their ability to teach science effectively.

STEBI-A for Science Teaching Self-Efficacy

To evaluate participants' science teaching self-efficacy (STE) before and after their participation in Every Day STEM for All (EDSTEMA), the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) was administered. The STEBI-A consists of 25 items scored on a Likert scale. It measures overall STE as well as utilizes two subscales to probe aspects of STE: personal science teaching efficacy (PSTE) and science teaching outcome expectancy (STOE) (Enochs & Riggs, 1990). The PSTE subscale measures teachers' confidence in their ability to teach science. It includes items that reflect how confident teachers feel in their ability to implement effective science teaching strategies, engage students in learning science, and manage classroom dynamics related to science instruction. The STOE subscale assesses teachers' belief that effective science teaching will lead to positive student outcomes.

Reliability and Validity. The STEBI-A was selected to examine science teaching self-efficacy (STE) in this study because it is a well-established instrument with high internal consistency and demonstrates strong validity and reliability evidence for measuring elementary teachers' confidence in their ability to teach science effectively (Bleicher, 2004; Enochs & Riggs, 1990). The STEBI has two versions: Form A, designed for in-service teachers, and Form B, intended for preservice teachers. STEBI-A was the more appropriate choice for this study because it aligns with the target population of elementary educators participating in Every Day STEM for All (EDSTEMA).

The STEBI-A has been shown to demonstrate strong construct validity, as it is aligned with the theoretical concepts of teaching self-efficacy (TSE) and expected outcomes. Research shows that the instrument effectively differentiates between varying levels of TSE and their expectations for student outcomes (Enochs & Riggs, 1990). Additionally, the STEBI-A's criterion-related validity has been supported by its correlations with other established measures of TSE, such as the Teachers' Sense of Efficacy Scale. The instrument also exhibits solid internal consistency, with Cronbach's alpha reliability coefficients generally exceeding .80 (Bleicher, 2004; Enochs & Riggs, 1990), further confirming its reliability.

In terms of reliability, the STEBI-A consistently demonstrates high internal consistency across samples. Enochs and Riggs (1990) reported Cronbach's alpha coefficients of .92 for the personal science teaching efficacy (PSTE) subscale and .77 for the science teaching outcome expectancy (STOE) subscale. Similarly, Bleicher (2004) found alpha coefficients of .90 and .76 for these respective subscales, further confirming the instrument's reliability across different populations of elementary educators. Collectively, these findings provide robust evidence of the STEBI-A's reliability and validity for assessing STE among in-service teachers.

Administration. Treatment group (TG) participants completed the STEBI-A pretest on the first day of Every Day STEM for All (EDSTEMA), before participating in any professional development (PD) activities. The posttest was administered at the end of the final follow-up session, following the completion of all PD activities. The comparison group (CG) was provided a digital link to the pretest at approximately the same time as the TG and was given a two-week window to submit their responses. Likewise, they received the posttest link around the same time as the TG and had two weeks to complete the assessment.

Scoring. Participants completed the STEBI-A at both the beginning and end of the Every Day STEM for All (EDSTEMA) program (see Table 4 for a sample questionnaire). Responses were recorded on a five-point Likert scale: 1 – *strongly disagree*, 2 – *disagree*, 3 – *uncertain*, 4 – *agree*, 5 – *strongly agree*. To ensure consistency in scoring, negatively worded items (items 3, 6, 8, 10, 13, 17, 19–22, 24, and 25) were reverse-coded, meaning that responses of *strongly agree* were reassigned a value of 1, while *strongly disagree* was reassigned a value of 5 (Riggs, 1988). The STEBI-A has demonstrated strong reliability and validity in prior studies, with reports of good internal consistency (Bleicher, 2004; Enochs & Riggs, 1990).

For each participant, item-level scores, overall average score, and changes between pretest and posttest administrations were calculated. Only participants who completed both the pretest and posttest were included in the final analysis. If a respondent did not complete both assessments, pairwise deletion was used to exclude the missing data from analysis. This approach was chosen over listwise deletion, as the missing scores did not impact other variables in the study. However, pairwise deletion assumes that data are missing at random (MAR). To assess this assumption, patterns of missing responses were examined to determine whether systematic issues could be responsible for the missing data.

Delimitations. Several decisions regarding the use of the STEBI-A instrument were intentional delimitations of this study. The study focused exclusively on elementary teachers participating in Every Day STEM for All (EDSTEMA) and comparison group (CG) teachers from the same partner districts to ensure comparability. The STEBI-A was selected rather than the preservice version (STEBI-B) because it was specifically developed for in-service teachers and aligns with the study's target population. The assessment was administered only at two points—before the summer institute (SI) and following the final follow-up session—to measure pre- and post-program changes in science teaching self-efficacy (STE). The analysis was limited to participants who completed both the pretest and posttest to maintain comparability across both groups. These delimitations were established to maintain coherence between the research purpose, data collection timeline, and program structure, while ensuring the feasibility of administration within the scope of EDSTEMA.

Limitations. Although the STEBI-A is a widely used and validated instrument, it is important to note that it relies on teacher self-reporting, which may introduce bias into the data. Teachers may unintentionally overestimate or underestimate their teaching efficacy due to personal perceptions, confidence levels, or external pressures such as evaluation systems or school culture. These factors can affect the accuracy of the results and complicate efforts to draw firm conclusions. Furthermore, while the STEBI-A has been successfully applied across a variety of educational contexts, its use in programs like Every Day STEM for All (EDSTEMA) may present additional limitations, as the unique goals and structures of such professional development (PD) initiatives could influence how teachers interpret and respond to survey items.

STEBI-A Summary. In summary, the STEBI-A provided a robust and well-validated measure of science teaching self-efficacy for the participants in this study. By capturing both personal science teaching efficacy (PSTE) and science teaching outcome expectancy (STOE), the instrument offered a nuanced picture of teachers' beliefs about their ability to teach science effectively. Its strong reliability and established use in prior research made it an appropriate choice for assessing changes in science teaching self-efficacy (STE) before and after Every Day STEM for All (EDSTEMA). While self-reporting introduces the potential for bias, the STEBI-A nevertheless served as a practical and meaningful tool for evaluating shifts in teachers' confidence and outcome expectancies, thereby contributing valuable data for understanding the impact of the program. Building on these teacher-level measures, the next section turns to student outcomes by examining Oklahoma Core Curriculum Test (OCCT) scores as an indicator of science achievement.

Oklahoma Core Curriculum Test for Student Achievement

To assess changes in science student academic achievement (SAA), fifth-grade science Oklahoma Core Curriculum Test (OCCT) scores were collected for the year prior to (Year 0) and the year following (Year 1) the Every Day STEM for All (EDSTEMA) summer institute (SI). Since a fourth-grade science OCCT did not exist, changes in science SAA for fourth-grade teachers' classes could not be measured. While alternative assessments such as district-level benchmarks or formative assessments were considered, they were not included in this study due to inconsistencies in administration and scoring across districts.

Reliability and Validity. The Oklahoma Core Curriculum Test (OCCT) was a statewide, standards-based assessment administered by the Oklahoma State Department of Education (OSDE) to measure student mastery of the Priority Academic Student Skills (PASS). According to the Oklahoma Technical Advisory Committee (TAC), the OCCT "met peer-review requirements for valid, reliable, and fair assessments" and was judged to be a high-quality instrument based on its comprehensive coverage of standards, measurement of the full range of student achievement, and the psychometric procedures used to ensure technical quality (Oklahoma Technical Advisory Committee, 2014). Although numerical reliability coefficients (e.g., Cronbach's α) for the fifth-grade science test were not publicly reported, the TAC's review indicates that the OCCT assessments were carefully evaluated to ensure their validity and reliability for accountability purposes.

During the study period, the state transitioned to a revised version of the science OCCT. At the time of data collection, psychometric information for the new version had not yet been released, so specific reliability and validity data for the updated assessment were unavailable. This transition presents a limitation, as potential differences in test structure, content emphasis, or scaling between Year 0 and Year 1 may affect score comparability. Despite this, the OCCT remained the most consistent and standards-aligned statewide measure of science achievement available during the study period, providing a practical means of examining changes in student academic achievement (SAA) among participating teachers' students.

Data Collection and Scoring. Because the Oklahoma Core Curriculum Test (OCCT) science assessment is administered only to fifth-grade students at the elementary level, only the scores from fifth-grade teachers in the treatment group (TG) were analyzed. Teachers self-reported their Year 0 scores at the beginning of Every Day STEM for All (EDSTEMA) and their Year 1 scores following the program's conclusion, after the state released results. Although self-reporting introduces the potential for reporting error or bias, efforts were made to mitigate these risks by cross-referencing available district records whenever possible.

To evaluate changes in science student academic achievement (SAA), overall pass rates for each year were calculated, and the difference in pass rates between Year 0 and Year 1 was determined. Only data sets containing scores for both years were included in the analysis to ensure completeness and comparability. This approach minimized bias from missing data and provided a more accurate measure of student achievement growth across the program period.

Delimitations. Several decisions regarding the use of Oklahoma Core Curriculum Test (OCCT) data were intentional delimitations of this study. The analysis was restricted to fifth-grade science scores because this is the only elementary grade level in which a state science

OCCT was administered. Data were further limited to teachers in the Every Day STEM for All (EDSTEMA) treatment group (TG) and their corresponding fifth-grade students, as no equivalent state assessment existed for fourth grade. Alternative measures of student academic achievement (SAA), such as district or classroom-level assessments, were excluded due to inconsistent administration and scoring practices across districts. Additionally, only teachers who provided complete OCCT data for both Year 0 and Year 1 were included in the analysis to ensure data integrity. These delimitations were designed to maintain consistency between the research purpose, available data, and the structure of the EDSTEMA program.

Limitations. Several limitations accompany the use of the Oklahoma Core Curriculum Test (OCCT) data. The most significant is the state's transition to a revised version of the science assessment between Year 0 and Year 1, for which psychometric documentation was not yet available. As a result, differences in test design, content alignment, or scoring procedures may have influenced comparability across years. In addition, the reliance on teacher self-reported scores, although verified where possible, introduces the potential for reporting error or bias. Finally, because the OCCT is only administered in fifth grade, the analysis excludes data from other elementary grade levels, limiting generalizability beyond the tested population. These limitations should be considered when interpreting changes in student academic achievement (SAA) as indicators of Every Day STEM for All's (EDSTEMA) program impact.

Oklahoma Core Curriculum Test Summary. In summary, the use of fifth-grade OCCT scores provided a standardized, state-reviewed measure of student science achievement aligned with the Priority Academic Student Skills (PASS). Despite limitations associated with the transition to a revised assessment and the use of self-reported data, the inclusion of both pre- and post-program scores enabled meaningful comparisons aligned with the timing of Every Day

STEM for All (EDSTEMA) implementation. While caution is warranted in interpretation, the OCCT data nevertheless offered valuable insight into potential program effects on student learning outcomes.

Instruments Summary

Taken together, the MOSART, STEBI-A, and OCCT provided a comprehensive set of measures to evaluate the impact of Every Day STEM for All (EDSTEMA) on teachers and students. The MOSART assessed changes in teachers' science content understanding (SCK) across domains, while the STEBI-A measured shifts in science teaching self-efficacy (STE), including both personal science teaching efficacy (PSTE) and science teaching outcome expectancy (STOE). The OCCT offered a standardized indicator of student academic achievement (SAA), capturing potential program effects at the classroom level. Although each instrument carried limitations—such as MOSART's original design for K–12 students, the self-reported nature of STEBI-A, and the state's transition to a new OCCT—their combined use provided a balanced and multidimensional picture of program outcomes. With these measures established, the following analysis section details the statistical procedures used to examine the relationships between EDSTEMA participation, teacher outcomes, and student achievement. These instruments are summarized in Table 7.

Table 7

Summary of Instruments Used in the Study

	MOSART	STEBI-A	OCCT
Purpose/ Variable Measured	Measured teachers' SCK and identification of misconceptions across four domains: space, life, physical, and Earth sciences.	25-item Likert-scale instrument measuring teachers' STE.	Measured students' science SAA to evaluate changes before and after EDSTEMA.
Participants/ Data Source	Elementary teachers in the TG and CG.	Elementary teachers in the TG and CG.	Fifth-grade students of participating TG teachers.
Administration & Timing	Administered as a pretest on the first day of the EDSTEMA SI and as a posttest at the end of the final follow-up session.	Administered pre- and post-EDSTEMA program.	Year 0 (pre-EDSTEMA) and Year 1 (post-EDSTEMA) state testing cycles.
Reliability and Validity Evidence	Reported *KR*–20 reliability = .73–.85.	Reliability coefficients: PSTE α = .90–.92, STOE α = .76–.77.	Oklahoma Technical Advisory Committee reported that the OCCT met peer-review standards for validity, reliability, and fairness.
Delimitations & Limitations	Delimitations: 34 selected items from 5–8 and 9–12 forms; four domains only; limited to EDSTEMA and CG teachers.	Delimitations: Form A selected for in-service teachers; data from complete pre- and posttests only.	Delimitations: Fifth-grade scores only; Year 0 and Year 1 data from TG teachers; excluded district-level tests.
	Limitations: Not validated for adults; originally designed for K–12 students.	Limitations: Self-report measure; potential response bias influenced by perceptions or school culture.	Limitations: State transition to revised OCCT; lack of psychometric data for new form; reliance on self-reported scores.

Note: MOSART = Misconceptions-Oriented Standards-Based Assessment Resources for Teachers; STEBI-A = Science Teaching Efficacy Beliefs Instrument – Form A; OCCT = Oklahoma Core Curriculum Test – Fifth-Grade Science; SCK = science content knowledge; STE = science teaching self-efficacy; SAA = student academic achievement; EDSTEMA = Every Day STEM for All; TG = treatment group; CG = comparison group; SI = summer institute; KR = Kuder-Richardson coefficient of reliability; PSTE = personal science teaching efficacy; STOE = science teaching outcome expectancy.

Analysis

Prior to engaging in Every Day STEM for All (EDSTEMA), the 44 participants in the treatment group (TG) completed the science content understanding (SCK) assessment and

Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) pretests. They completed these assessments again as posttests at the conclusion of EDSTEMA. The comparison group (CG) also completed these assessments before and after EDSTEMA. After EDSTEMA concluded, the SCK assessment and STEBI-A results were analyzed using descriptive and inferential statistics to compare pretest and posttest scores, as well as TG and CG results. The type of test used depended upon whether the data were normally distributed, as determined using Shapiro–Wilk tests, and whether homogeneity of variances existed, as determined using Levene's test. In addition to teacher-level outcomes, student performance data were also examined. Fifth-grade science members of the TG provided their students' Oklahoma Core Curriculum Test (OCCT) pass rates for the two consecutive years before and after EDSTEMA.

All scores were analyzed statistically using a combination of software tools. Microsoft Excel was used to calculate descriptive statistics and conduct initial inferential tests; G*Power was used for post hoc power analyses; and AI–Therapy Statistics (2025) was used to verify Excel results and generate graphics.

Science Content Knowledge

The first research question (RQ1) of this study focused on whether participation in Every Day STEM for All (EDSTEMA) would significantly improve teachers' science content understanding (SCK). To measure this, an SCK assessment consisting of selected Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) items (outlined in Appendix F) was administered at the beginning of the summer institute (SI) and again at the conclusion of the year-long EDSTEMA program. For each set of data, descriptive statistics (*M*, *SD*, ranges) were first computed to provide an overview of the data distribution and central tendencies. Each individual's SCK assessments were scored according to the percentage of questions they answered correctly. Higher scores indicated higher levels of SCK. The posttest results were compared to the pretest results using a paired-samples *t*-test, and the treatment group (TG) results were compared to the comparison group (CG) results using a Welch's *t*-test.

Treatment Group Pretest Versus Posttest

A paired-samples *t*-test was conducted to determine whether the difference in pretest and posttest means within the treatment group (TG) was statistically significant. This test was appropriate because the same group of individuals completed the same assessment before and after the intervention (Denis, 2016).

Prior to conducting the *t*-test, its assumptions were evaluated. The assumption of normality was assessed using Shapiro–Wilk tests and evaluated via visual inspection of histograms. For paired-samples *t*-tests, the normality of the difference scores was assessed using Shapiro–Wilk tests and visual inspection of histograms. The assumption of homogeneity of variances was tested using Levene's test.

To ensure adequate statistical power, a power analysis was conducted. Because participant recruitment was limited to the number of teachers enrolled in the program, an a priori power analysis was not feasible. Instead, a post hoc power analysis was conducted to evaluate the achieved power. Effect size was calculated using Cohen's *d* to determine the magnitude of the observed difference.

Treatment Group Versus Comparison Group Pretests

To assess whether significant differences in science content understanding (SCK) existed between the treatment group (TG) and comparison group (CG) prior to the start of Every Day STEM for All (EDSTEMA), a Welch's *t*-test was performed on the pretest scores of both groups.

This test was chosen because it does not assume equal variances, making it appropriate given the disparity in sample sizes (Derrick et al., 2016). Levene's test was conducted to check the equality of variances.

Despite the unequal sample sizes, Welch's *t*-test was preferred due to its robustness in handling unequal variances and sample sizes, because it is robust to unequal variances and unequal sample sizes histograms. The assumption of normality was confirmed through Shapiro–Wilk tests, and further supported via histograms. The effect size was calculated using the formula for Cohen's *d* for Welch's *t*-tests, and statistical power was assessed post hoc.

Treatment Group Versus Comparison Group Posttests

Given the unequal sample sizes and unequal variances between the treatment group (TG) and comparison group (CG), a Welch's *t*-test was again selected to compare the two groups' posttest scores, as it does not assume equal variances and is more appropriate under these conditions (Derrick et al., 2016). The assumption of normality was assessed using the Shapiro–Wilk test and evaluated via visual inspection of histograms. To ensure adequate statistical power, a post hoc power analysis was conducted. Descriptive statistics were computed to provide an overview of the data distribution and central tendencies.

Change From Pretest to Posttest

The change in science content understanding (SCK) scores from pretest to posttest for both groups was calculated by subtracting pretest scores from posttest scores for each participant. An independent-samples *t*-test compared the gain scores (posttest minus pretest) of the treatment group (TG) and comparison group (CG) to determine whether changes in SCK were significantly greater for TG. This test assumes that the difference in scores from pretest to posttest is normally distributed, which was confirmed using Shapiro–Wilk tests and evaluated via visual inspection of histograms. To assess the magnitude of the observed effect, Cohen's *d* was calculated. Additionally, to ensure sufficient statistical power, a post hoc power analysis was conducted.

Science Content Knowledge Analysis Summary

In summary, the analyses of science content understanding (SCK) demonstrated meaningful changes associated with participation in Every Day STEM for All (EDSTEMA). Paired-samples t-tests showed whether individual teachers improved from pretest to posttest, while Welch's t-tests provided a robust comparison between the treatment group (TG) and comparison group (CG) despite unequal variances and sample sizes. Effect sizes and post hoc power analyses further contextualized the magnitude and strength of the observed differences, offering a fuller understanding of program impact. Taken together, these findings establish a foundation for examining whether growth in teachers' SCK was accompanied by shifts in their confidence to teach science effectively. The next section therefore turns to results from the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A), which measured changes in science teaching self-efficacy (STE).

Science Teaching Self-Efficacy

The second research question (RQ2) of this study examined whether participation in Every Day STEM for All (EDSTEMA) would improve science teaching self-efficacy (STE). To measure this, the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) questionnaire was administered at the beginning of the summer institute (SI) and again at the conclusion of the year-long EDSTEMA program. For each participant, the average score for each item and the overall average score were calculated. Descriptive statistics were calculated for each data set to summarize the key features of the data. A variety of tests were performed to compare the means of pretest and posttest scores, as well as the scores of the treatment group (TG) and comparison group (CG). Specifically, a Wilcoxon signed-rank test was used to compare the TG's pretest and posttest scores, while an analysis of covariance (ANCOVA) was conducted to compare the posttest scores of the TG and CG, controlling for any baseline differences.

Treatment Group Pretest Versus Posttest

Because the normality assumption was violated for the paired *t*-test, a Wilcoxon signed-rank test was conducted as a nonparametric alternative to determine whether the difference in pretest and posttest means was statistically significant. This test was selected as a nonparametric alternative to the paired *t*-test because a Shapiro–Wilk analysis found that posttest scores were not normally distributed (Derrick et al., 2016). Effect size was calculated using rank biserial correlation to determine the magnitude of the observed difference. To ensure adequate statistical power, a post hoc power analysis was conducted.

Treatment Group Versus Comparison Group

To assess the effectiveness of the Every Day STEM for All (EDSTEMA) intervention, an analysis of covariance (ANCOVA) was conducted with posttest STEBI-A scores as the DV and pretest STEBI-A scores as the covariate, to compare the posttest scores of the treatment group (TG) and comparison group (CG), while controlling for potential differences in pretest scores. This approach allowed for the control of pre-existing group differences, ensuring that any observed differences in posttest scores could be attributed to EDSTEMA itself rather than any pretest variability among the groups (Tabachnick & Fidell, 2013). By adjusting for the pretest scores, ANCOVA provided a more accurate comparison of the groups' outcomes after the intervention.

Science Teaching Self-Efficacy Analysis Summary

Taken together, the STEBI-A results provided valuable evidence about how Every Day STEM for All (EDSTEMA) influenced teachers' confidence in their ability to teach science effectively. The findings highlighted meaningful changes in teachers' self-perceptions following participation in the program. These shifts are important because research has consistently shown that teacher self-efficacy is strongly linked to instructional quality and student learning outcomes. Building on this foundation, the next section extends the analysis beyond teacher beliefs to explore whether these improvements translated into measurable gains for students. Specifically, student academic achievement (SAA) will be examined using fifth-grade Oklahoma Core Curriculum Test (OCCT) scores, providing an additional perspective on EDSTEMA's overall impact.

Student Science Achievement

To investigate the final research question (RQ3), this study examined whether a teacher's participation in Every Day STEM for All (EDSTEMA) would improve their science student academic achievement (SAA). Teachers provided their fifth-grade science Oklahoma Core Curriculum Test (OCCT) pass rates for two consecutive years: the year prior to participating in EDSTEMA (Year 0) and the year during the program (Year 1). The data from these two years were compared to identify any differences in science SAA following the teachers' participation in EDSTEMA.

To assess whether significant differences existed between the two years' pass rates, a Welch's *t*-test was performed. This test was selected because Levene's test indicated unequal variances, and Welch's *t*-test does not assume equal variances, making it well-suited for the unequal sample sizes in this dataset (Derrick et al., 2016). The assumption of normality was evaluated using Shapiro–Wilk tests and supported by histogram inspection. Effect size was calculated using Cohen's d for Welch's *t*-tests, and statistical power was assessed post hoc to evaluate the adequacy of the analyses.

Student Science Achievement Summary

In sum, the use of fifth-grade OCCT scores provided a standardized means of assessing student science achievement before and after Every Day STEM for All (EDSTEMA) participation. This approach ensured that student-level outcomes were examined alongside teacher-level measures. Together, the MOSART, STEBI-A, and OCCT created a comprehensive framework for evaluating program impacts. The next chapter presents the results of these analyses, beginning with teacher outcomes and moving to student achievement.

Summary

This study investigates the impact of the Every Day STEM for All (EDSTEMA) professional development (PD) program on elementary teachers' science content understanding (SCK), science teaching self-efficacy (STE), as well as their students' student academic achievement (SAA) in science. The program was developed in response to growing concern about the decline in scientific literacy across the United States—a trend often linked to reduced instructional time for science in the early grades and elementary teachers' limited confidence in teaching the subject. EDSTEMA aimed to address these issues by providing sustained, content-focused PD that emphasized both science knowledge and pedagogy.

University faculty skilled in both domains facilitated the program, modeling lessons through the 5E model of science instruction. These lessons were designed to engage teachers in active learning, enhancing their understanding of science content while simultaneously developing their ability to apply that knowledge in the classroom. The program's structure emphasized hands-on experiences and the application of science teaching strategies in real-world contexts.

The study employed a quasi-experimental pretest–posttest design to evaluate the program's effectiveness. The study employed a treatment group (TG) that participated in EDSTEMA and a comparison group (CG) that did not. Because the treatment and comparison groups were of unequal sizes, tests that assume equal variances could have been affected. To mitigate this risk, the analysis employed statistical procedures robust to unequal sample sizes and variances—for instance, using Welch's t-tests instead of independent-samples t-tests and applying analysis of covariance (ANCOVA) to control for baseline differences.

The participants were elementary math and science teachers from five suburban school districts within the Oklahoma City metropolitan area. Data were collected using a selection of MOSART questions to measure SCK, the STEBI-A to measure science teaching self-efficacy (STE), and OCCT scores to measure SAA in science. Pretest and posttest scores were analyzed using descriptive and inferential statistics, including paired-samples *t*-tests and nonparametric alternatives, to assess changes in SCK, STE, and SAA. Additionally, analyses such as *t*-tests and ANCOVAs were conducted to compare TG and CG scores.

The analysis of EDSTEMA's impact on teachers' SCK, STE, and students' SAA contributes valuable insights to the field of science education. Given the limited research examining PD programs that integrate science content and pedagogy, the findings offer important evidence of how such initiatives can foster teacher confidence and improve student outcomes. These results may inform the design of future PD programs and policy decisions aimed at strengthening science instruction in elementary schools.

Chapter 3 outlined the research design, population and sample, EDSTEMA program structure, instruments, and data analysis procedures used to examine the effects of the EDSTEMA professional development program on teachers' science content knowledge (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) in science. Building on these methodological foundations, Chapter 4 presents the results of the statistical analyses conducted to address each of the three research questions. The chapter is organized by outcome variable—SCK, STE, and SAA—and reports descriptive statistics, inferential test results, effect sizes, and measures of statistical power to evaluate the impact of EDSTEMA on both teacher and student outcomes.

Chapter 4: Results

This chapter presents the results of a quasi-experimental study designed to examine the effect of participation in Every Day STEM for All (EDSTEMA)—a sustained, coherent, cohort-based professional development (PD) program designed to actively engage teachers in science content. The study employed a pretest–posttest design with nonequivalent groups to evaluate EDSTEMA's effects on both teacher and student outcomes. Specifically, the analysis addressed three research questions (RQs):

RQ1: Does participation in EDSTEMA significantly improve elementary teachers' science content knowledge (SCK)?

RQ2: Does participation in EDSTEMA significantly improve elementary teachers' science teaching self-efficacy (STE)?

RQ3: Does participation in EDSTEMA significantly improve student academic achievement (SAA) in science compared to the previous year?

The sections that follow describe the treatment group (TG) and comparison group (CG), summarize the data collection and analytical procedures, and present results aligned with each research question. A variety of statistical tests were used to examine within-group changes from pretest to posttest and to compare outcomes between the TG and CG; the rationale for each test is discussed in the relevant sections. Finally, the chapter summarizes key findings and explores their implications for elementary science education and professional development (PD).

Participants

This study was designed to assess the extent to which participation in Every Day STEM for All (EDSTEMA) improves teachers' science content understanding (SCK) and science teaching self-efficacy (STE) and their students' academic achievement (SAA) in science. The treatment group (TG) consisted of 44 teachers from 26 elementary schools across five districts. The comparison group (CG) included 44 teachers from 18 schools in those same districts. All participants in both groups completed pre- and posttest administrations of the SCK assessment and/or the Science Teaching Efficacy Beliefs Instrument Form A (STEBI-A).

Efforts were made to recruit CG members with characteristics closely matching those of the TG, including grade level, teaching assignment, and years of experience. The following sections describe these relevant characteristics in detail (see Table 2 for a summary).

Treatment Group

Most participants (91%) were female, representing 26 public schools across five districts. Of these, 34 teachers (77%) worked in low-income schools, defined as those where at least half of the students qualified for free or reduced-price lunch (FRL).

All participants taught at least one math or science course at the fourth- or fifth-grade level. Specifically, 21 (48%) taught fourth grade, and 27 (61%) taught fifth grade. One participant taught both fourth and fifth grades, so percentages may exceed 100%. Subject assignments varied: four (9%) taught only science, 31 (70%) taught science plus other subjects, and 5 (11%) taught only non-science subjects.

Teaching experience ranged from one to 30 years (*M* = 11, *Mdn* = 8.5). Two participants (4.5%) were first-year teachers, 13 (30%) had two to five years, eight (18%) had 6–10 years, 10 (23%) had 11–15 years, six (14%) had 16–20 years, three (7%) had 21–25 years, and two (5%) had 26–30 years.

The treatment group (TG) members were not randomly selected. To participate in Every Day STEM for All (EDSTEMA), teachers had to apply, commit to participate in all 86 hr of planned professional development (PD), and agree to develop and teach a lesson based on what they learned during the program. In exchange, they received a stipend of $2,000, resource books, and classroom materials.

Because the program required significant personal time during summer and weekends, some self-selection bias may have occurred, potentially resulting in an overrepresentation of highly motivated teachers in the TG. However, the substantial monetary incentive may have mitigated this bias by attracting teachers who otherwise would not have applied.

Comparison Group

The comparison group (CG) did not participate in Every Day STEM for All (EDSTEMA). Instead, they continued their usual professional development (PD) activities. This design helps isolate the effect of EDSTEMA by providing a comparison to typical PD experiences. Similar to the treatment group (TG), the majority (96%) of the 44 CG members were female teachers. They were recruited from 18 public schools in the same five districts as the TG. Nineteen participants (73%) worked in low-income schools where at least half of the student body qualified for free or reduced lunch (FRL).

Only teachers who taught at least one math or science course at the fourth- or fifth-grade level were eligible for the CG. Thirteen participants (50%) taught fourth grade, 15 participants (58%) taught fifth grade, and three participants taught both grades, so they are included in both counts. Percentages may total more than 100% because some teachers taught multiple grades. Twenty-two participants (85%) taught science in addition to one or more other subjects, while four participants (15%) taught only subjects other than science.

The participants' teaching experience ranged from one to 30 years (M = 10.5, Mdn = 9.5). Specifically, three (12%) were first-year teachers, seven (27%) had two to five years, two (9%)

had six to 10 years, seven (27%) had 11–15 years, four (15%) had 16–20 years, one (4%) had 21–25 years, and two (8%) had 26–30 years.

The CG members were not selected randomly from the partner schools. Teachers were required to submit applications and agree to complete all required assessments. In return, CG members were offered a stipend of $100 for each assessment they completed. As a result, self-selection bias may exist, meaning the CG may not fully represent the broader population. To mitigate this potential bias, efforts were made to match CG members as closely as possible to the TG in terms of school characteristics, personal demographics, and teaching experience.

Attrition

Although Every Day STEM for All (EDSTEMA) initially recruited strong participation from both the treatment group (TG) and comparison group (CG), some attrition occurred over the course of the study. Participation was voluntary, and group assignment was based on teachers' decisions to enroll in the professional development (PD) program or continue with their typical professional learning (PL) experiences.

A total of 44 teachers were recruited for the TG and 44 for the CG. All 44 TG participants completed the summer institute (SI) and both pre- and post-assessments for the science content understanding (SCK) measure, resulting in a 0% attrition rate. This full participation likely reflects the structured nature of the PD program and the completion-based financial incentives.

In contrast, the CG experienced substantial attrition for the SCK and Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) measures. Of the 44 teachers initially recruited for the CG, only 17 completed both the SCK pretest and posttest. For the STEBI-A, 17 completed the pretest, and 44 completed the posttest. The discrepancy between pre- and posttest participation reflects a procedural issue identified mid-study that prevented retroactive administration of the pretest.

For the Oklahoma Core Curriculum Test (OCCT) in science, data were obtained from all 27 fifth-grade science teachers in the TG. This high completion rate was expected, given that OCCT data were drawn from existing state testing systems.

Table 8 summarizes participant recruitment and completion across measures for both groups. A more detailed discussion of the causes and implications of attrition is provided in the Limitations section.

Participants Summary

In summary, both the treatment group (TG) and comparison group (CG) consisted of elementary teachers with similar demographic and professional backgrounds. However, self-selection and differences in incentives may have influenced participation. The TG included teachers who applied to and fully participated in the Every Day STEM for All (EDSTEMA) program, while the CG continued with their usual professional development (PD) activities.

Although some attrition occurred, the final samples remained sufficiently comparable to allow for meaningful analysis of EDSTEMA's impact on teacher and student outcomes. These characteristics and conditions establish the foundation for the subsequent discussion of data collection procedures to measure changes in science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) in science.

The following section describes the instruments and procedures used to gather data from both groups, including the administration of pretests and posttests, teacher surveys, and student achievement measures.

Table 8

Attrition Summary

	Science Content Knowledge Assessment (MOSART)		Science Teaching Self-Efficacy Assessment (STEBI-A)		Student Academic Achievement Assessment (OCCT)	
Group	TG	CG	TG	CG	TG	CG
Recruited	44	44	44	44	27	—
Completed Pretest	44	17	44	17	27	—
Completed Posttest	44	38	40	38	27	—
Completed Both	44	17	40	17	27	—
Attrition Rate	0%	61%	9	61%	0%	—
Notes	Full completion; posttest administered at SI conclusion	Delayed recognition of non-completed pretests; not administered retroactively	Four did not complete posttest due to scheduling conflicts	Delayed recognition of non-completed pretests; not administered retroactively	Complete data; drawn from state test records	Not obtained due to oversight in data collection

Note. TG = treatment group; CG = comparison group. Attrition rate calculated as the percentage of participants who did not complete both pretest and posttest for each measure.

Data Collection

To evaluate whether participation in Every Day STEM for All (EDSTEMA) improved teachers’ science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) in science, several types of data were collected. Before engaging in EDSTEMA, the 44 participants in the treatment group (TG) completed two pretests. One consisted of a sampling of Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) questions designed to measure the TG’s initial SCK on concepts covered in the EDSTEMA summer institute (SI). The other pretest was the full version of the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) questionnaire, used to assess the TG’s initial level of STE. At the conclusion of EDSTEMA, both assessments were administered to the TG again to determine any changes in SCK and STE. These assessments were administered concurrently to the comparison group (CG) to enable comparative analysis. The results were then analyzed to determine whether participation in EDSTEMA led to greater improvements in SCK and STE compared to nonparticipation.

In addition, TG members provided their science SAA scores for the two years surrounding their participation in EDSTEMA. Specifically, participants self-reported their fifth-grade science Oklahoma Core Curriculum Test (OCCT) pass rates for the year prior to EDSTEMA (Year 0) and for the program year (Year 1). Scores were analyzed to explore whether EDSTEMA participation influenced students' science SAA. While OCCT scores were self-reported, efforts were made to verify accuracy by cross-referencing with available school or district records when possible.

In summary, the study combined teacher and student data to assess EDSTEMA's impact. Pre- and posttests of SCK and STE were administered to both groups, while fifth-grade science OCCT results provided student-level outcomes. Together, these measures created a robust dataset for analyzing changes in knowledge, self-efficacy, and achievement.

Data Analysis

To evaluate whether participation in Every Day STEM for All (EDSTEMA) improved teachers' science content knowledge (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA), several analyses were conducted. Descriptive statistics summarized pretest and posttest scores, and inferential tests—paired-samples *t*-tests, independent-samples *t*-tests, Welch's *t*-tests, and analyses of covariance (ANCOVA)—were used to compare results within and between groups.

Additionally, the treatment group's (TG's) Oklahoma Core Curriculum Test (OCCT) pass rate scores from Year 0 were compared to their scores from Year 1 to assess the effect of EDSTEMA on science SAA. The processes used to analyze each set of data—as well as the results of these analyses—are discussed in the following sections and summarized in Table 9.

Science Content Knowledge Results

The first research question (RQ1) of this study examined whether participation in Every Day STEM for All (EDSTEMA) improved teachers' science content understanding (SCK). To measure this outcome, an SCK assessment consisting of a selection of Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) items (outlined in Appendix F) was administered at the beginning of the summer institute (SI) and again at the conclusion of the year-long EDSTEMA program. The posttest scores were compared with the pretest scores to evaluate within-group growth, while treatment group (TG) results were also compared with comparison group (CG) results to assess program impact. Together, these analyses provided a basis for determining the extent to which EDSTEMA participation influenced teachers' SCK. The following subsection describes the scoring procedures used for the SCK assessment.

Table 9

Summary of Statistical Tests Used for Each Research Question

RQ & Instrument	Comparison	Statistical Test Used	Rationale
Effect of EDSTEMA on SAA (OCCT)	Year 0 versus Year 1	Welch's *t*-test	Assumption of equal variances violated

Note. RQ = research question; SCK = science content knowledge; MOSART = Misconceptions-Oriented Standards-Based Assessment Resources for Teachers; TG = treatment group; CG = comparison group; *n* = sample size; EDSTEMA = Every Day STEM for All; STE = science teaching self-efficacy; STEBI-A = Science Teaching Efficacy Beliefs–Form A; ANCOVA = analysis of covariance; SAA = student academic achievement; OCCT = Oklahoma Core Curriculum Test; Year 0 = year prior to participation in EDSTEMA; Year 1 = year of participation in EDSTEMA.

Scoring

The science content understanding (SCK) assessment consisted of 20 multiple-choice items selected from various Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) tests, representing the content covered during the summer institute (SI). Each correct response was scored as 1, and incorrect responses were scored as 0. The change in average score from pretest to posttest was also calculated for each respondent. A summary of the percentage of correct responses for each item for the pretest and posttest, along with the difference in scores between the two administrations, is presented in Appendix G.

Of the 44 members of the comparison group (CG), 27 (61%) did not complete the pretest, and six (14%) did not complete the posttest. To address the missing data, a *pairwise deletion* approach was used, excluding participants with missing pretest and/or posttest scores on the SCK assessment. The missing scores were considered missing at random (MAR), as their absence was unrelated to other factors in the study (Little and Rubin, 1987, as cited in Shadish et al., 2002). This approach was preferred over *listwise deletion*, as it minimized the loss of data without introducing *imputation bias* (Tabachnick & Fidell, 2013) and did not impact other variables in this study (Shadish et al., 2002).

As a result, 61% of the data for this variable were missing or excluded, which represented a significant limitation. Missing data of this magnitude substantially reduced and unbalanced sample sizes across analyses, decreased statistical power, and increased the likelihood of Type I or Type II errors, thereby potentially skewing findings.

Analysis

Having established the structure and scoring procedures of the science content understanding (SCK) assessment, the next step was to analyze the results to determine whether participation in Every Day STEM for All (EDSTEMA) produced measurable gains. The following section presents the statistical analyses conducted to compare pretest and posttest scores within the treatment group (TG), as well as between the TG and comparison group (CG). These analyses provide insight into whether EDSTEMA participation was associated with significant improvements in teachers' SCK.

All 44 members of the TG completed both the SCK assessment pretest and posttest. Seventeen members of the CG completed the pretest and 24 completed the posttest. Only participants with both pretest and posttest scores were included in the analysis. Descriptive statistics were calculated to summarize each data set. The TG posttest results were compared to the pretest results using a paired-samples *t*-test, and the TG results were compared to the CG results using a Welch's *t*-test. The change in scores from pretest to posttest in both groups was analyzed using an independent-samples *t*-test. The following sections describe the analyses that were conducted and discuss the findings.

Treatment Group Pretest Versus Posttest. The treatment group (TG) ($n = 44$) was administered the science content understanding (SCK) assessment as a pretest and a posttest. Descriptive statistics (see Appendix H) for the pretest indicated a mean score of $M = 59.09$ ($SD = 14.48$), and for the posttest, the mean score was $M = 74.77$ ($SD = 10.56$). To assess the normality of the data, a Shapiro–Wilk test was conducted for both the pretest and posttest scores. The results indicated that the data were normally distributed for both the pretest, $W(44) = 0.96$, $p = .17$, and the posttest, $W(44) = 0.96$, $p = .12$. Additionally, histograms of the TG's SCK assessment pretest and posttest scores, illustrating normal distribution, are shown in Figures M1 and M2. Levene's test for equality of variances confirmed that the assumption of equal variances was not violated, $F(43) = 3.47$, $p = .07$.

Analysis of item-level performance revealed that the percentage of correct responses increased on 15 items, remained unchanged for one item, and decreased on four items (see Appendix G). The decrease in scores for certain items may indicate that these topics were more complex or less well-understood by participants. The items showing decreased performance addressed high school-level science concepts and were written at a level requiring higher-order cognitive processing, such as application, analysis, or evaluation. Due to the difficulty of these items, participants may have responded by guessing randomly on both the pretest and posttest. As a result, fluctuations in scores on these items may not reflect true changes in understanding but rather the variability associated with low-confidence or uninformed guesses. Figure M3 illustrates the shift in mean score from pretest to posttest.

A paired-samples *t*-test was conducted to examine the difference in mean scores from pretest to posttest in the TG. This test was selected given that the same group of participants completed the same assessment before and after the intervention, the data were normally distributed, and the assumption of equal variances was met (Denis, 2016). The results revealed a statistically significant difference, $t(43) = -10.55$, $p < .001$, indicating that participation in Every Day STEM for All (EDSTEMA) led to a substantial 15-point increase (59.09 to 74.77) in scores from the pretest to the posttest. The effect size, as measured by Cohen's *d*, was very large ($d = 1.24$), suggesting a strong treatment effect (Hattie, 2009). An effect size of this magnitude corresponds to an improvement of more than one full standard deviation, suggesting that the intervention led to meaningful improvements in participants' SCK.

To ensure adequate statistical power, a power analysis was conducted. The analysis indicated that with an $n = 44$, power ≈1.00 (100%) to detect an effect size of 1.24 at an alpha level of 0.05. This high level of power suggests that the study is highly unlikely to miss any meaningful effects, supporting the robustness of the findings. These results suggest that teachers who participated in the program are likely better equipped to deliver accurate science instruction, potentially leading to improved student outcomes.

Treatment Group Versus Comparison Group. Participation in Every Day STEM for All (EDSTEMA) was voluntary, which may limit the generalizability of the findings. Self-selection bias could be present, as participants may have exhibited greater levels of motivation than nonparticipants (Shadish et al., 2002). In practical terms, this bias may have resulted in a participant pool composed disproportionately of teachers who were already inclined toward professional growth, more confident in their ability to benefit from science-focused professional development (PD), or more willing to invest personal time in exchange for long-term instructional improvement. These teachers may have also had stronger existing support systems or fewer competing responsibilities outside of school.

As a result, the program's observed effects might not be replicable among a broader, more representative group of teachers, particularly those with lower intrinsic motivation, limited time, or less prior exposure to science content. This limitation is especially important when considering the scalability of EDSTEMA, as future implementation in diverse school settings may not yield similar outcomes without addressing variations in teacher readiness or willingness to engage.

To mitigate self-selection bias, efforts were made to ensure that the comparison group (CG) had characteristics similar to those of the treatment group (TG), particularly in terms of prior science content understanding (SCK), years of teaching experience, and demographic composition (see Table 2 for a detailed comparison).

Statistical analyses confirmed that there were no significant differences between the TG and the CG on key variables at baseline, thus supporting the comparability of the groups. This suggests that any observed posttest differences are more likely due to the intervention rather than preexisting differences. Descriptive statistics for the pretest and posttest SCK assessment administrations for both the TG and CG are provided in Appendix H.

Treatment Group Versus Comparison Group Pretests. To determine whether significant differences in science content understanding (SCK) existed between the treatment group (TG) ($n = 44$) and comparison group (CG) ($n = 17$) prior to the start of Every Day STEM for All (EDSTEMA), a Welch's *t*-test was conducted on the pretest scores of both groups. This test was chosen because it does not assume equal variances (Derrick et al., 2016), making it appropriate given the disparity in sample sizes. Levene's test for equality of variances indicated no significant difference in variances, $F(1, 55) = 1.36$, $p = .25$, confirming that the assumption of homogeneity of variance was not violated. However, given the unequal sample sizes, Welch's *t*-test was still used as a more robust approach. Normality was confirmed by Shapiro–Wilk tests on the TG, $W(40) = 0.96$, $p = .17$, and CG, $W(17) = 0.95$, $p = .39$. Additionally, histograms of the TG and CG pretest scores illustrate normal distributions, as shown in Figures M1 and M4.

The results of the Welch's *t*-test revealed no significant difference between the TG pretest scores ($M = 59.09$, $SD = 14.48$) and CG pretest scores ($M = 54.71$, $SD = 19.00$), $t(24) = 0.87$, $p = .39$. The effect size, as measured by Cohen's *d*, was large ($d = 0.87$) (Cohen, 1988). Statistical power was calculated post hoc and determined to be 0.82. This indicates that there was an 18% chance that a significant effect was not detected by this analysis.

Although the TG ($n = 44$) was larger than the CG ($n = 17$), Welch's *t*-test mitigated the risks associated with unequal sample sizes. These results indicated that both the TG and CG began with comparable levels of SCK before EDSTEMA began, supporting the assumption that both groups began the study with comparable levels of science knowledge. As a result, the CG was confirmed to be a suitable comparison group for evaluating the impact of the intervention. However, although no significant differences were found at pretest, factors such as prior training, instructional quality, or access to resources may still have influenced outcomes.

Treatment Group Versus Comparison Group Posttests. The results from the Welch's *t*-test indicate that both the treatment group (TG) and the comparison group (CG) started with comparable levels of science content understanding (SCK) before Every Day STEM for All (EDSTEMA) began. This baseline similarity ensures that any improvements observed in the TG after the intervention can be more confidently attributed to EDSTEMA's impact rather than preexisting differences between the groups. Therefore, a comparison of the two groups' posttest scores was conducted to assess the effect of EDSTEMA on SCK.

To assess the normality of the data, a Shapiro–Wilk test was conducted for both the TG and CG and posttest scores. The results indicated that the data were normally distributed for both the TG, $W(44) = 0.96$, $p = .12$, and the posttest, $W(44) = 0.91$, $p = .10$. Additionally, histograms of the TG's SCK assessment pretest and posttest scores, illustrating normal distribution, are shown in Figures M1 and M2. Levene's test for equality of variances indicated a significant difference in variances between the two groups, $F(1,55) = 7.10$, $p = .01$, suggesting that the assumption of equal variances was violated. Given this and the unequal sample sizes (TG: n = 44, CG: n = 17), a Welch's *t*-test was chosen as it does not assume equal variances and is more appropriate under these conditions (Derrick et al., 2016).

Results from the SCK assessment posttest showed that the TG (M = 74.77, SD = 10.56) outperformed the CG (M = 54.12, SD = 10.56) by an average of 20 percentage points, $t(22)$ =

4.92, $p < .001$. These findings indicate that participation in EDSTEMA led to a significant increase in teacher SCK compared to nonparticipants. The effect size, as calculated using Cohen's *d* formula for Welch's *t*-tests was $d = 1.43$, indicating a very large effect (Hattie, 2009).

To ensure adequate statistical power, a post hoc power analysis was conducted. The analysis indicated that with unequal sample sizes between the TG ($n = 44$) and CG ($n = 17$), the study had 100% power (1.00) to detect an effect size of $d = 1.43$ at an alpha level of .05. This indicates that the study was highly powered to detect the observed effect, providing strong confidence in the reliability of the result.

Overall, these results highlight the substantial benefits of the EDSTEMA program in enhancing SCK. The results indicate that teachers who participated in EDSTEMA significantly outperformed those who did not, demonstrating the program's strong influence on SCK. This suggests that EDSTEMA provided substantial learning benefits, providing teachers with greater science content knowledge than those who did not receive the training.

Change From Pretest to Posttest. The change in science content understanding (SCK) scores from pretest to posttest for both the treatment group (TG) ($n = 44$) and comparison group (CG) ($n = 17$) was calculated by subtracting pretest scores from posttest scores. To assess the normality of the data, a Shapiro–Wilk test was conducted for both the TG and CG change scores. The results indicated that the data were normally distributed for both the TG, $W(44) = 0.93$, $p = .24$, and the CG, $W(44) = 0.96$, $p = .11$. Histograms illustrating the normal distribution in the change in SCK assessment scores for the TG and CG are shown in Figures M6 and M7. Levene's test for equality of variances indicated no significant difference in variances between the two groups, $F(1,55) = 0.43$, $p = .51$.

An independent-samples *t*-test was then conducted to examine the difference in TG and CG change scores. This test was selected because the data were normally distributed and the assumption of equal variances was met (Denis, 2016). The results revealed a statistically significant difference, $t(59) = -5.53$, $p < .001$. These results indicate that teachers who participated in Every Day STEM for All (EDSTEMA) ($M = 15.93$, $SD = 10$) experienced significantly greater growth in SCK compared to those who did not participate ($M = 0.59$, $SD = 11$). Descriptive statistics for the change in TG and CG scores between pretest and posttest administrations are provided in Appendix H. The effect size, calculated using Cohen's *d*, is $d = 1.53$, indicating a large effect (Cohen, 1988). Given that effect sizes above 0.40 are considered educationally meaningful (Hattie, 2009), these results suggest that participation in EDSTEMA led to substantial learning gains.

To ensure adequate statistical power, a post hoc power analysis was conducted. The analysis indicated that with unequal sample sizes between the TG ($n = 44$) and CG ($n = 17$), the study had 100% power (1.00) to detect an effect size of 1.53 at an alpha level of 0.05. This means that the study is highly powered to detect such an effect, should it truly exist. However, while the observed effect size was large, future research should consider replication with a larger sample to confirm these findings.

Overall, these results demonstrated that teachers who completed the program experienced a 16-percentage point increase in SCK on average, while those who did not participate showed no meaningful change. The results provide compelling evidence that high-quality, targeted professional development (PD) programs like EDSTEMA can lead to significant knowledge gains.

Given the stark difference in knowledge growth between participants and nonparticipants, expanding access to programs like EDSTEMA could have widespread benefits for science education. However, as participation in the program was voluntary, self-selection bias remains a consideration. Future research could be strengthened by implementing a randomized controlled trial or improving the matching of CG to TG to account for differences between participants and nonparticipants.

Alternate Tests Considered. Because the Every Day STEM for All (EDSTEMA) study involved a pretest–posttest design with both treatment and comparison groups, an analysis of covariance (ANCOVA) was initially considered to determine whether posttest differences between groups remained significant when statistically controlling for pretest performance. ANCOVA is frequently used in professional development studies to increase statistical precision, as it adjusts posttest means by accounting for variability explained by the pretest covariate (Tabachnick & Fidell, 2013). This made ANCOVA an attractive option for comparing teachers' science content understanding (SCK) outcomes while controlling for baseline differences.

The ANCOVA results indicated that both the covariate and group membership were significant predictors of posttest performance. Specifically, pretest scores strongly predicted posttest outcomes, $F(1, 59) = 137.76$, $p < .001$, and group membership continued to account for significant variance in posttest scores after controlling for pretest, $F(1, 59) = 84.56$, $p < .001$. However, the critical assumption of homogeneity of regression slopes was not satisfied, $F(1, 58) = 7.16$, $p = .010$. This violation undermines the interpretability of the adjusted group comparison because it indicates that ANCOVA does not equate groups on the covariate in a consistent manner. In other words, the covariate (pretest) influenced posttest outcomes differently depending on group membership, making ANCOVA an inappropriate choice as the primary analytic method in this study.

Given this violation, independent-samples *t*-tests were determined to be a more appropriate analytic approach. The *t*-tests, conducted on gain scores and posttest outcomes, do not rely on the homogeneity of slopes assumption and provide a more straightforward evaluation of whether the treatment and comparison groups differed. The findings from the ANCOVA supported the general conclusion that the treatment group (TG) outperformed the comparison group (CG); however, because the model assumptions were not fully satisfied, *t*-tests were prioritized as the main analyses reported. This decision ensured that results were both statistically sound and aligned with the methodological rigor expected for this study.

Science Content Knowledge Summary

The results from multiple analytic approaches—which are summarized in Table 10—demonstrated that participation in Every Day STEM for All (EDSTEMA) produced significant gains in teachers' science content understanding (SCK). Teachers in the treatment group (TG) not only improved substantially from pretest to posttest but also outperformed their comparison group (CG) peers by a wide margin, with large effect sizes providing strong evidence of program impact. These findings underscore the power of sustained, content-focused professional development (PD) to deepen teachers' SCK, which is a critical foundation for high-quality instruction. While SCK is essential, teacher confidence in their ability to effectively deliver science instruction also plays a vital role in shaping classroom practice. The following section examines whether EDSTEMA participation influenced teachers' STE, addressing the second research question of this study.

Science Teaching Self-Efficacy

The second research question (RQ2) of this study examined whether participation in Every Day STEM for All (EDSTEMA) would improve teachers' science teaching self-efficacy (STE). To measure this, the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) was administered to participants at the beginning of the summer institute (SI) and again at the conclusion of the year-long EDSTEMA program. This instrument provided a measure of teachers' confidence in their ability to effectively teach science, as well as their beliefs about the potential impact of their instruction on student learning.

The posttest results for the treatment group (TG) were compared to their pretest results to assess growth over time, and they were also compared to the responses from the comparison group to evaluate differences attributable to program participation. These analyses were designed to determine whether EDSTEMA participation was associated with meaningful improvements in STE, an important factor in effective science instruction.

Scoring

The STEBI-A questionnaire, which uses a Likert scale to measure science teaching self-efficacy (STE), was administered to participants at both the beginning and end of the Every Day STEM for All (EDSTEMA) program (see Table 4 for a sample questionnaire). Each response was assigned a numerical value: 1 = *strongly disagree*, 2 = *disagree*, 3 = *uncertain*, 4 = *agree*, 5 = *strongly agree*. To ensure consistency in scoring, negatively worded items (items 3, 6, 8, 10, 13, 17, 19–22, 24, and 25) were reverse-coded, meaning that *strongly agree* was reassigned a value of 1, while *strongly disagree* was reassigned a value of 5 (Riggs, 1988).

Table 10

Summary of Statistical Results for Science Content Knowledge

Source	Descriptive Stats	Normality	Variance	Test	Effect size	Power
TG MOSART Pretest x Posttest Results						
TG Pretest	*n* = 44 *M* = 59.09 *SD* = 14.48	Yes- Shapiro–Wilk *W* = .96, *p* = .17	Levene's *F* = 3.47, *p* = .07 Equal	Paired samples *t*-test *t*(43)=-10.55, *p* < .001 Significant difference	Cohen's *d* for independent samples = 1.24 Very large	1.00
TG Posttest	*n* = 44 *M* = 74.77 *SD* = 10.56	Yes- Shapiro–Wilk *W* = .96, *p* = .12				
TG x CG MOSART Pretests Results						
TG Pretest	*n* = 44 *M* = 59.09 *SD* = 14.48	Yes- Shapiro–Wilk *W* = .96, *p* = .17	Levene's *F* = 1.36, *p* = .25	Welch's *t*-test *t*(24)=.87, *p*=.39	Cohen's *d* for Welch's = .87 Large	.82
CG Pretest	*n* = 17 *M* = 54.71 *SD* = 19	Yes- Shapiro–Wilk *W* = .95, *p* = .39	Equal	No significant difference		
TG x CG MOSART Posttests Results						
TG Posttest	*n* = 44 *M* = 74.77 *SD* = 10.56	Yes- Shapiro–Wilk *W* = .96, *p* = .12	Levene's *F* = 7.10, *p* = .01	Welch's *t*-test *t*(22)=4.92, *p*<.001	Cohen's *d* for Welch's = 1.43 Very large	1.00
CG Posttest	*n* = 17 *M* = 54.12 *SD* = 10.56	Yes- Shapiro–Wilk *W* = 0.91, *p* =.10	Unequal	Significant difference		
Change in MOSART Scores						
TG Change	*n* = 44 *M* = 15.93 *SD* = 10	Yes- Shapiro–Wilk *W* = 0.93, *p* =.24	Levene's *F* = 0.43, *p* = .51	Independent samples *t*-test *t*(59)=-5.53, *p*<.001	Cohen's *d* for independent samples *t*-tests = 1.53 Large	1.00
CG Change	*n* = 17 *M* = 0.59 *SD* = 11	Yes- Shapiro–Wilk *W* = 0.96, *p* =.11	Equal	Significant difference		

Note: Change is defined as posttest score minus pretest score. MOSART = Misconceptions-Oriented Standards-Based Assessment Resources for Teachers; TG = treatment group; *n* = sample size; *M* = mean; *SD* = standard deviation; *W* = Shapiro–Wilk test statistic; *p* = probability value; *F* = F test statistic; *t* = *t*-test statistic; CG = comparison group.

This approach follows the scoring procedures established by Riggs and Enochs (1990) to ensure consistency in interpreting the construct. An examination of the reverse-coded items indicated that, once recoded, they tracked in the same general direction as the straightforward items. However, the gains on several of these items were smaller and somewhat noisier, which is a common occurrence in survey instruments that include negatively phrased statements. In this study, the reverse-coded items still reflected the same underlying construct of science teaching self-efficacy as the positively worded items, though with slightly less precision.

For each participant, an average score was calculated for each item, along with an overall composite score and a growth score between the pretest and posttest administrations. Change scores were computed by subtracting pretest scores from posttest scores, thereby providing an individual-level measure of growth in STE. These calculations allowed the study to capture not only general trends in self-efficacy across groups but also the variation in how individual teachers responded to the EDSTEMA program. A full summary of these results, including item-level and overall changes, is presented in Appendix I.

Missing data reduced the number of participants available for analysis. Four members of the treatment group (TG) did not complete both the pretest and posttest, which prevented the calculation of growth scores for those teachers. In the comparison group (CG), 27 members did not complete the pretest, leaving only 17 out of 44 with usable pretest–posttest data, although all 44 completed the posttest. These uneven completion rates limit direct comparisons between groups and highlight a potential source of bias, as participants who completed all assessments may differ systematically from those who did not.

Despite these limitations, the available data provided a sufficient sample to conduct statistical analyses of both within-group growth and between-group differences in STE. The following section presents the analyses performed on these data, highlighting whether participation in EDSTEMA was associated with measurable improvements in teachers' self-efficacy in science instruction.

Analysis

Descriptive statistics were calculated for each data set to summarize the key features of the data. A variety of tests were performed to compare the means of pretest and posttest scores, as well as the scores of the treatment group (TG) and comparison group (CG). Specifically, a Wilcoxon signed-rank test was used to compare the TG's pretest and posttest scores, as this nonparametric test does not assume normality and was appropriate given the distributional properties of the data. In addition, an analysis of covariance (ANCOVA) was conducted to compare the posttest scores of the TG and CG, controlling for any baseline differences.

A total of forty TG members completed both the STEBI-A pretest and posttest. Seventeen out of 44 CG members also completed the pretest, and all 44 completed the posttest. To address the 27 missing CG scores, a pairwise deletion approach was employed, excluding all participants who did not complete the STEBI-A pretest. Missing data were considered to be missing at random (MAR) because their absence was unrelated to other factors in the study (Little and Rubin, 1987, as cited in Shadish et al., 2002). This method of pairwise deletion was chosen over listwise deletion because the deleted scores had no impact on other variables in this study, and this method allowed for the preservation of as much data as possible. As a result, 61% of the CG data for this variable were either missing or excluded.

This significant amount of missing data represents a limitation of the study, as it could introduce bias into the study, resulting in reduced and inconsistent sample sizes across analyses, and potentially decreasing statistical power and skewing findings (Shadish et al., 2002). If participants who completed the pretest differed in meaningful ways (e.g., motivation, experience, teaching effectiveness) from those who did not, the findings may not fully represent the broader CG population.

Two STEBI Constructs. The STEBI-A is designed to measure two related but distinct constructs that stem from Bandura's (1977) social cognitive theory: Personal Science Teaching Efficacy Beliefs (PSTE) and Science Teaching Outcome Expectancy (STOE) (Riggs & Enochs, 1990). PSTE reflects a teacher's confidence in their own ability to effectively teach science—for example, their belief that they understand science content well enough to explain it to students or can answer students' science questions. In contrast, STOE measures the extent to which teachers believe that their efforts in teaching science will translate into student learning gains. In other words, PSTE focuses on the teacher's self-perception of competence, while STOE addresses the perceived link between teaching actions and student outcomes. Prior research suggests that professional development (PD) is more likely to influence PSTE than STOE, because teachers can directly experience changes in their own knowledge and skills, whereas their beliefs about broader causal relationships between teaching and learning tend to be more stable.

As shown in Appendix J, the data reveal a distinction between teachers' PSTE and their STOE. Teachers in the treatment group (TG) showed stronger growth in PSTE, with average scores increasing from 3.65 at pretest to 4.01 at posttest (a gain of 0.35 points). By contrast, their STOE increased more modestly, from 3.40 to 3.72 (a gain of 0.32 points). Teachers in the comparison group (CG) showed little change in either construct, with a small increase in PSTE (a gain of 0.12 points) and virtually no change in outcome expectancy (a small gain of 0.03 points). These results align with Bandura's (1977) theory that PD is more likely to directly influence teachers' confidence in their own instructional skills than to substantially shift their broader beliefs about the relationship between teaching and student achievement.

Treatment Group Pretest Versus Posttest. A Wilcoxon signed-rank test was conducted to determine if the difference between the treatment group's (TG) pretest and posttest scores was significant. This test was selected as a nonparametric alternative to the paired *t*-test because, although the pretest scores were normally distributed ($W = 0.98$, $p = .85$), as confirmed by a Shapiro–Wilk test (Figure M8), the posttest scores were not ($W = 0.94$, $p = .05$) (Derrick et al., 2016). Given the borderline significance of the p value, skewness and kurtosis were calculated, and a Q–Q plot was generated. The skewness was found to be -.13, $p = .71$, suggesting the data's skew was similar to a normal distribution. However, the kurtosis value was -1.07, $p = .02$, indicating *platykurtic* distribution, meaning the data had a flatter peak and shorter tails compared to a normal distribution (Shadish et al., 2002). The Q–Q plot (see Figure M9) further confirmed the deviation from normality. Given these findings, a Wilcoxon signed-rank test was chosen as a nonparametric alternative to the paired *t*-test (Derrick et al., 2016).

The Wilcoxon signed-rank test revealed that the posttest scores ($M = 96.03$, $SD = 8.75$) were significantly higher than the pretest scores ($M = 88.30$, $SD = 8.50$), ($Z = 5.25$, $p < .001$, $r = .95$). Appendix K provides descriptive statistics for the pretest and posttest scores. Of the participants, five showed a decrease in scores, while 35 showed an increase. Figure M10 illustrates the shift in mean score from pretest to posttest.

These results suggest that participation in Every Day STEM for All (EDSTEMA) leads to an improvement in science teaching self-efficacy (STE). The effect size, measured using rank-biserial correlation was $r = .95$, indicating a large effect (Cohen, 1988). To ensure adequate statistical power, a power analysis was conducted. The analysis indicated that with $n = 40$, the study had 100% power (1.00) to detect an effect size of .95 at an alpha level of .05. This means that the study is highly powered to detect such an effect, should it truly exist.

Teachers' posttest scores were significantly higher than their pretest scores, indicating that EDSTEMA had a strong positive effect on their STE. A majority of participants (35 out of 40) showed improvement, reinforcing EDSTEMA's effectiveness across different individuals. However, the presence of five participants with decreased scores warrants further investigation into possible individual differences in response to the program. Overall, the findings from this analysis suggest that high-quality, targeted professional development (PD) programs like EDSTEMA can significantly enhance teachers' self-efficacy in science instruction.

Treatment Group Versus Comparison Group. A STEBI-A pretest and posttest were administered to both the treatment group (TG) ($n = 40$) and comparison group (CG) ($n = 17$). These assessments were conducted to measure changes in participants' science teaching self-efficacy (STE) before and after Every Day STEM for All (EDSTEMA). Descriptive statistics for the pretest and posttest scores are provided in Appendix K. For the TG, the pretest mean was $M = 88.30$ ($SD = 8.50$), and the posttest mean was $M = 96.03$ ($SD = 8.75$). For the CG, the pretest mean was $M = 87.76$ ($SD = 9.29$), and the posttest mean was $M = 88.29$ ($SD = 7.89$).

To compare the posttest scores between the TG and CG while adjusting for pretest scores, an analysis of covariance (ANCOVA) was conducted. The results revealed a significant difference in mean change scores between the groups, $F(1, 54) = 16.15$, $p < .001$. This indicates that the TG showed a significantly greater improvement in scores compared to the CG. The η^2 value for this analysis was .23, indicating a large effect size (Cohen, 1988). This effect size suggests that approximately 23% of the variance in STEBI-A change scores can be explained by participation or nonparticipation in EDSTEMA. The power of the analysis was calculated to be

.90, indicating that the study was sufficiently powered to detect the effect. These results support the conclusion that EDSTEMA had a statistically significant impact on participants' scores.

The results demonstrate that EDSTEMA had a positive effect on STE, leading to significant improvements in confidence and teaching effectiveness in science instruction. Findings show that teachers who participated in EDSTEMA showed a significant increase in STE scores compared to those who did not participate. The average posttest scores in the TG were significantly higher than pretest scores, with 35 out of 40 participants improving. Teachers in the CG who did not receive the intervention showed little to no growth in STE over the same period. The ANCOVA effect size was large, indicating that the program accounted for a meaningful portion of the variance in self-efficacy outcomes. These findings suggest that high-quality, sustained professional development (PD) programs like EDSTEMA can significantly improve teacher self-efficacy, leading to better instructional outcomes.

Interpretation of Declines in STE Over Time

Although most participants experienced increases in STE, five teachers exhibited a slight decline (an average of 1.5 points), highlighting the need to investigate individual differences in response to the program rather than assuming uniformly positive trajectories for all participants.

These variations may stem from contextual factors such as teaching assignment, school-level support, or differences in prior content knowledge (CK). Another possibility is that participation in EDSTEMA created a degree of cognitive disequilibrium for some teachers (Piaget, 1977). As they engaged more deeply with rigorous science content and ambitious instructional strategies, some participants may have come to recognize previously unacknowledged gaps in their understanding or skills. From a constructivist perspective, this shift reflects a movement from assimilating new ideas into existing conceptions of "good" science teaching to accommodating those conceptions in light of more complex, discipline-specific expectations. During this accommodation process, teachers may temporarily feel less confident, not because they have regressed, but because they are developing a more accurate and critical appraisal of their practice.

Several design features of EDSTEMA could have contributed to this process of cognitive disequilibrium and recalibration. First, the strong emphasis on science content knowledge required teachers to confront challenging disciplinary ideas and common misconceptions. Engaging with diagnostic assessments and tasks that surfaced conceptual gaps may have made some participants acutely aware of the limits of their own understanding.

Second, the program's active learning structures—such as engaging in investigations as learners, analyzing student work, and designing and sharing lesson plans—rendered teachers' thinking public. In these contexts, participants were asked not only to absorb new information but also to make their reasoning visible and open to feedback.

Third, EDSTEMA's collaborative format created opportunities for social comparison. Working alongside colleagues who demonstrated greater comfort with scientific practices, or who appeared more adept at facilitating inquiry-based discussions, may have led some teachers to view their own practice in a less favorable light. This may help explain why a subset of teachers might report feeling less efficacious at the end of the program than at the beginning, despite having expanded their instructional repertoire.

This phenomenon aligns with the Dunning–Kruger effect, which suggests that individuals with limited knowledge may initially overestimate their competence, but as their knowledge deepens, they develop a more modest and calibrated self-assessment (Kruger & Dunning, 1999). In this context, a temporary decline in self-efficacy may not reflect a lack of learning, but rather a critical shift in self-awareness that marks the beginning of meaningful cognitive and professional growth. EDSTEMA's emphasis on challenging content, substantive conversations, and opportunities for comparison with peers may have intensified this process for some participants. Thus, small dips in STE for a subset of teachers could be interpreted as evidence of productive disequilibrium and increasing professional sophistication, rather than as an unintended negative consequence of the PD.

Science Teaching Self-Efficacy Summary

Overall, the findings, which are summarized in Table 11, indicate that participation in Every Day STEM for All (EDSTEMA) led to significant improvements in science teaching self-efficacy (STE). Both domains of the STEBI-A—personal science teaching efficacy (PSTE) and science teaching outcome expectancy (STOE)—showed meaningful gains for the treatment group (TG) compared to the comparison group (CG).

While the raw increases were similar in size (PSTE +0.35 vs. STOE +0.32), the CG showed almost no change, which means the relative impact of EDSTEMA was strong in both areas. This suggests that the program not only strengthened teachers' confidence in their own instructional abilities but also enhanced their belief that effective science teaching leads to student learning. Large effect sizes confirmed the robustness of these improvements. Since STE is closely linked to instructional practice and, ultimately, student learning, the next section examines whether these teacher-level gains translated into measurable improvements in student achievement, addressing the third research question.

Table 11

Summary of Statistical Results for Science Teaching Self-Efficacy

Source	Descriptive Stats	Normality	Variance	Test	Effect size	Power
TG STEBI-A Pretest x Posttest Results						
TG Pretest	n = 40 M = 88.30 SD = 8.50	Yes- Shapiro–Wilk W = 0.98, p = .85	Levene's F = .51, p = .48 Equal	Wilcoxon signed rank test Z = 5.25, p<.001 Significant difference	Rank biserial correlation r = .95 Very large	1.00
TG Posttest	n = 40 M = 96.03 SD = 8.75	Borderline- Shapiro–Wilk W =.94, p = .05				
STEBI-A ANCOVA Results						
TG Pretest	n = 40 M = 88.30 SD = 8.50	Yes- Shapiro–Wilk W = 0.98, p = .85	Yes Levene's F = .63, p = .43 Equal	F(1,111)=16.15, p<.001 Significant difference	η^2_p = .23 Large	.72
TG Posttest	n = 40 M = 96.03 SD = 8.75	Borderline- Shapiro–Wilk W =.94, p = .05				
CG Pretest	n = 17 M = 87.76 SD = 9.29	Yes- Shapiro–Wilk W =.99, p = 1.00				
CG Posttest	n = 17 M = 88.29 SD = 7.89	Yes- Shapiro–Wilk W =.97, p = .80				

Note: STEBI-A = Science Teaching Efficacy Beliefs Inventory–Form A; TG = treatment group; CG = comparison group; n = sample size; M = mean; SD = standard deviation; W = Shapiro–Wilk test statistic; p = probability value; F = F test statistic; Z = standardized Wilcoxon test statistic; r = correlation coefficient; η^2_p = partial eta squared coefficient.

Student Science Achievement

To investigate the final research question (RQ3), this study examined whether a teacher's participation in Every Day STEM for All (EDSTEMA) was associated with improvements in student academic achievement (SAA) in science. Specifically, teachers provided their fifth-grade science Oklahoma Core Curriculum Test (OCCT) pass rates for two consecutive years: the year prior to their participation in EDSTEMA (Year 0) and the year during the program (Year 1). These data allowed for a direct comparison of student performance before and after teachers' involvement in the program, offering insight into the potential transfer of professional development (PD) effects to classroom outcomes. A Welch's t-test was conducted to compare the Year 0 and Year 1 pass rates, enabling the study to determine whether significant differences in student science achievement could be attributed to teachers' participation in EDSTEMA.

Scoring

Because the OCCT state science assessment is administered only to fifth-grade students at the elementary level, only the scores from fifth-grade science teachers in the treatment group (TG) were included in this analysis. To ensure consistency and comparability, scores were analyzed only when both Year 0 (the year prior to Every Day STEM for All (EDSTEMA) participation) and Year 1 (the year of participation) data were available for the same teacher. For each teacher, overall pass rates were calculated for both years, and the change in pass rates from Year 0 to Year 1 was computed to capture growth over time. These calculations provided a basis for determining whether participation in EDSTEMA was associated with measurable improvements in student science achievement on the state assessment.

Analysis

A Welch's *t*-test was conducted in order to determine if there were differences in how the classes taught by treatment group (TG) members performed on the OCCT state science assessment before and after participating in Every Day STEM for All (EDSTEMA). This test was chosen because, although the OCCT exam was administered to a different set of students in both years, the focus of this study is on the impact of the teacher, and it seeks to determine how their students' performance changed after participating in EDSTEMA. Further, a Levene's test revealed unequal variances between the two groups, $F = 0.98$, $p = .33$. Shapiro–Wilk tests confirmed that scores are normally distributed for both Year 0 ($W = 0.96$, $p = .45$) and Year 1 ($W = 0.93$, $p = .09$), as shown in Figure M13 and M14 (Denis, 2016).

The results from Year 0 ($M = 70.15$, $SD = 18.78$) and Year 1 ($M = 79.41$, $SD = 15.80$) showed that teacher participation in EDSTEMA led to significantly increased OCCT scores, $t(51) = -1.96$, $p = .03$. Descriptive statistics for each data set are provided in Appendix L. This statistically significant improvement suggests that participation in EDSTEMA had a meaningful impact on teachers' instructional effectiveness, as reflected in their students' assessment performance. The effect size, as measured by Cohen's *d*, was $d = 0.59$, indicating a moderate effect (Cohen, 1988). This suggests that the program contributed to meaningful gains in teacher instructional effectiveness as reflected in student achievement outcomes.

Although these findings are statistically significant and practically meaningful, the power analysis revealed that the study had only 31% power (.31) to detect an effect size of this magnitude at an alpha level of .05. This relatively low level of power reflects the small sample size ($n = 27$) and underscores the need for caution when interpreting the results. With such limited power, there is an increased risk of Type II error. A concise overview of the statistical tests, results, and effect sizes is provided in Table 12.

Practical Implications

Overall, these findings provide strong evidence that sustained, coherent, and cohort-based professional development (PD) programs, such as Every Day STEM for All (EDSTEMA), are highly effective in improving both teacher and student outcomes. This study demonstrated that teachers who participated in EDSTEMA showed significant gains in science content understanding (SCK) and science teaching self-efficacy (STE), which in turn contributed to higher scores in science student academic achievement (SAA). The results also highlight the importance of active engagement in subject-specific content and the benefits of long-term, structured PD over shorter, less comprehensive training.

Teachers can apply the knowledge gained from such PD programs by integrating stronger science content expertise into their instruction, using evidence-based teaching strategies with greater confidence, and fostering a more engaging learning environment for students. Enhanced STE can lead to more effective classroom practices, better student engagement, and improved learning outcomes in science.

Table 12

Summary of Statistical Results for Student Academic Achievement

Difference in OCCT Scores from Year 0 to Year 1 Results						
Source	Descriptive Stats	Normality	Equal Variance	Test	Effect size	Power
Year 0	*n* = 27 *M* = 70.15 *SD* = 18.78	Yes Shapiro–Wilk *W* = .96, *p* = .45	No- Levene's *F* = 0.98 *p* = .33	Welch's *t*-test *t*(51)=-1.96, *p*=.03 Significant differences	Cohen's *d* for Welch's *d* = 0.59 Moderate	0.31
Year 1	*n* = 27 *M* = 79.41 *SD* = 15.80	Yes Shapiro–Wilk *W* = .93, *p* = .09				

Note: Change is defined as posttest score minus pretest score. OCCT = Oklahoma Core Curriculum Test; Year 0 = year prior to Every Day STEM for All (EDSTEMA); Year 1 = year of EDSTEMA; *n* = sample size; *M* = mean; *SD* = standard deviation; *W* = Shapiro–Wilk test statistic; *p* = probability value; *F* = F test statistic; *t* = *t*-test statistic.

For school districts, these findings emphasize the value of investing in sustained, content-focused PD programs that promote teacher learning for SAA. Policies should prioritize long-term, cohort-based PD models that actively engage teachers in subject-specific content and provide ongoing support. Additionally, districts should implement mechanisms to assess the effectiveness of PD initiatives to ensure they lead to measurable improvements in teacher performance and student success.

In conclusion, this study affirms that thoughtfully designed PD programs like EDSTEMA can serve as a powerful lever for strengthening elementary science education. By enhancing teachers' SCK, boosting their STE, and supporting gains in SAA, EDSTEMA demonstrates the transformative potential of sustained, coherent, and content-rich professional learning. These findings contribute to the broader PD literature while offering actionable insights for educators, administrators, and policymakers committed to improving science teaching and learning.

Ultimately, expanding access to high-quality PD models of this kind holds promise for fostering more confident teachers, more engaged students, and a stronger culture of scientific literacy in schools.

Conclusion

This chapter presented a detailed analysis of the quasi-experimental study evaluating the impact of Every Day STEM for All (EDSTEMA), a sustained, coherent, cohort-based professional development (PD) program focused on science content knowledge, on elementary teachers and their students. The findings provide compelling evidence that participation in EDSTEMA significantly improved teachers' science content understanding (SCK), science teaching self-efficacy (STE), as well as their students' academic achievement (SAA). While the results are promising, several limitations—including self-selection bias, missing comparison group data, and the use of self-reported SAA scores—necessitate caution when interpreting these findings.

Despite these limitations, the comparability of the treatment group (TG) and comparison group (CG) at baseline and statistically significant gains observed at the conclusion of the PD suggest that EDSTEMA effectively supports teacher learning and positively influences student outcomes. The study highlights the importance of long-term, content-rich PD approaches that actively engage teachers in deepening their subject knowledge and confidence in teaching.

Further implications of these findings, alongside comprehensive recommendations for future research to address study limitations and explore the sustained effects of EDSTEMA, will be discussed in Chapter 5.

Chapter 5: Discussion

This study examined the outcomes associated with the Every Day STEM for All (EDSTEMA) professional development (PD) program, which was funded by the Math and Science Partnership (MSP) initiative and designed to improve science education by providing elementary teachers with research-based, sustained, and collaborative PD. Specifically, the study investigated how participation in EDSTEMA was related to changes in teachers' science content knowledge (SCK), science teaching self-efficacy (STE), and students' academic achievement (SAA) in science. This research addressed a gap in the literature by exploring how such PD models are associated with both teacher growth and student learning over time. The findings are intended to inform the development of future science PD programs by illustrating the potential of sustained, cohort-based approaches to support teacher effectiveness and, in turn, promote stronger student outcomes in science education.

This chapter presents a comprehensive discussion of the findings from this quasi-experimental study, which examined outcomes associated with participation in the EDSTEMA PD program for elementary science teachers' SCK, STE, and their students' SAA in science. It begins with a summary of results, followed by an interpretation of the findings. The discussion highlights areas of alignment and divergence from existing literature and illustrates the study's contribution to the broader field of science education and PD research. Theoretical and practical implications are addressed, particularly in terms of their significance for strengthening teacher PD models and instructional practices in science classrooms. The chapter also discusses study limitations to provide context for interpreting the results and concludes with recommendations for future research and final conclusions drawn from the study as a whole.

Summary of Results

The purpose of this study was to examine the impact of the Every Day STEM for All (EDSTEMA) professional development (PD) program on teacher and student outcomes by addressing the following three research questions:

RQ1: Does participation in EDSTEMA significantly improve elementary teachers' science content knowledge (SCK)?

RQ2: Does participation in EDSTEMA significantly improve elementary teachers' science teaching self-efficacy (STE)?

RQ3: Does participation in EDSTEMA significantly improve student academic achievement (SAA) in science compared to the previous year?

This section presents a review of the research findings and provides a summary of the results corresponding to each research question.

Impact of EDSTEMA on Science Content Knowledge

This study examined whether participation in Every Day STEM for All (EDSTEMA) was associated with significant increases in elementary teachers' science content understanding (SCK). The findings are consistent with the hypothesis (H1), which stated that teachers who participated in EDSTEMA would demonstrate a statistically significant increase in SCK from pretest to posttest compared to those who did not participate. Teachers in the treatment group (TG) showed a significant gain in SCK from pretest to posttest, with a large effect size indicating meaningful improvement. Additionally, when compared to the comparison group (CG), the TG demonstrated significantly higher posttest scores and greater overall growth in SCK. These results were supported by strong statistical power and similar baseline scores between groups, which strengthens confidence that participation in EDSTEMA was related to the observed gains, while recognizing that the quasi-experimental design limits strong causal claims.

Impact of EDSTEMA on Science Teaching Self-Efficacy

This study also examined whether participation in Every Day STEM for All (EDSTEMA) was associated with significant increases in elementary teachers' science teaching self-efficacy (STE). The findings support the hypothesis (H2), which stated that teachers who participate in EDSTEMA would demonstrate a statistically significant increase in STE from pretest to posttest compared to those who did not participate. Teachers in the treatment group (TG) showed a significant gain in STE from pretest to posttest, with a small but practical effect size indicating a meaningful improvement. In practical terms, even modest gains in self-efficacy are often associated with noticeable changes in classroom practice, such as a teacher's willingness to attempt new instructional strategies or to persist when faced with student misconceptions. These incremental changes, though statistically small, may accumulate over time to produce meaningful shifts in instructional quality and student engagement.

Additionally, when compared to the comparison group (CG), the TG demonstrated significantly higher posttest scores and greater overall growth in STE. Although the between-group effect size was small, the direction and consistency of the results provide evidence of a positive pattern associated with participation in EDSTEMA. This suggests that the program may have provided teachers with experiences that reinforced their confidence in teaching science, even if the measurable effect was not dramatic. It also raises the possibility that the most substantial implications of STE gains may be indirect—emerging more clearly in long-term instructional practices and student outcomes rather than in immediate posttest scores. These findings were supported by adequate statistical power and comparable baseline scores between groups, which strengthens confidence that participation in EDSTEMA was related to the observed gains, while acknowledging that the quasi-experimental design limits strong causal conclusions.

Impact of EDSTEMA on Student Academic Achievement in Science

This study also examined whether participation in Every Day STEM for All (EDSTEMA) was associated with changes in student academic achievement (SAA) in science. The findings support the hypothesis (H3), which stated that students taught by teachers who participated in EDSTEMA would demonstrate statistically higher SAA scores compared to students taught by those same teachers in the previous year. In this study, one cohort of students was assessed before their teacher participated in EDSTEMA, and a second cohort was assessed after the same teacher completed the program; following the intervention year, the latter cohort showed substantially higher average scores on the state science assessment, with a large effect size indicating a meaningful difference in academic performance. Although the comparison was made across two different student cohorts, the focus on teachers' participation in EDSTEMA revealed that higher student scores occurred in the year after teachers completed the program. These results were supported by strong statistical power and a normal distribution of scores, which strengthens confidence in the robustness of the finding that the post-EDSTEMA cohort scored higher in SAA than the pre-EDSTEMA cohort, even though the quasi-experimental design and cohort differences prevent definitive causal conclusions about the specific effects of EDSTEMA on teacher practice or student achievement.

Interpretation of Results

The findings across all three research questions provide compelling evidence that participation in Every Day STEM for All (EDSTEMA) was associated with strengthened teacher science content understanding (SCK), enhanced science teaching self-efficacy (STE), and higher student academic achievement (SAA) in science. Each outcome demonstrated statistically significant gains, supported by adequate statistical power and consistent with the program's design as a sustained, coherent, and content-focused professional development model. While the magnitude of effects varied, with particularly strong gains in SCK and more modest but meaningful growth in STE, the overall pattern suggests the potential value of long-term, structured professional development (PD) for supporting both teacher learning and student outcomes.

This section interprets these results in the context of existing literature, drawing connections to prior studies and theoretical frameworks in order to more fully understand the patterns observed in EDSTEMA and their implications for science education. It provides an interpretation of the results in relation to the three research questions that guided the study. Each set of findings is examined alongside existing literature to consider the extent to which the outcomes align with, extend, or diverge from previous research on professional development in science education. Through this comparison, the discussion highlights how EDSTEMA can inform the broader understanding of effective PD, particularly in relation to strengthening teacher SCK, supporting STE, and promoting higher SAA. By situating the results within the larger body of research, this chapter aims to provide a more nuanced understanding of the program's role and potential influence and to identify implications for both practice and future research.

Science Content Knowledge

The findings from this study provide strong evidence of substantial gains in elementary teachers' SCK among those who participated in the EDSTEMA PD program. Within the treatment group (TG), there was a substantial increase in mean scores from pretest (M = 59.09) to posttest (M = 74.77), a 15-point gain supported by a very large effect size (d = 1.24). This magnitude of effect indicates that the changes were not only statistically significant but also educationally meaningful, consistent with Hattie's (2009) threshold for impactful learning interventions. Such results are consistent with the interpretation that EDSTEMA supported teachers in developing deeper conceptual understanding of science content, thereby strengthening their foundation for classroom instruction. Moreover, the consistency of these gains across participants suggests that improvements in SCK were widespread within the TG rather than confined to a small subset of highly motivated teachers.

Item-level analysis provided additional insights into learning trends. While most items showed improvement on the posttest, a few declined—particularly those aligned with high school–level standards and requiring higher levels of cognitive processing. This pattern may reflect item difficulty or guessing behaviors, indicating that variability in scores might represent uncertainty rather than a regression in conceptual understanding. In some cases, a decline in scores may signal increased self-awareness and a more critical, realistic appraisal of current practice rather than an actual loss of competence. As teachers encounter more rigorous science content and ambitious instructional approaches, they may begin to see science teaching as more complex and demanding than they previously realized, which can temporarily lower their confidence. This interpretation is consistent with research suggesting that deeper exposure to expert practice can move individuals from "unconscious incompetence" to "conscious incompetence," a stage in which they recognize the distance between their current practice and a more advanced vision of instruction. Thus, for some participants, lower posttest science teaching self-efficacy (STE) scores may reflect a productive stage of professional growth in which they are recalibrating their beliefs about their own effectiveness in light of new learning.

When comparing the TG to the comparison group (CG), which had similar baseline characteristics, TG participants scored 20 percentage points higher on the posttest, with a very large between-group effect size ($d = 1.43$). This likely reflects the deliberate alignment between EDSTEMA content and the SCK assessment. The SCK assessment used in this study was specially designed to measure the key science concepts emphasized in the PD, drawing on the principles of backward design (Wiggins & McTighe, 2005). First, the EDSTEMA team identified the specific SCK learning outcomes they wanted teachers to achieve; second, they developed assessment items that would provide evidence of those outcomes; and only then did they design the instructional sequence and learning experiences to help participants meet those targets. By structuring the program in this way, the professional learning experiences, assessment tasks, and desired content outcomes were tightly aligned, supporting the interpretation that measured gains in SCK reflect meaningful growth in the focal concepts of the PD.

TG participants, who engaged deeply in activities designed to teach the targeted content during the PD, were expected to perform significantly better on the posttest than on the pretest when they had not yet received the instruction. In contrast, CG participants, who received no exposure to the EDSTEMA content, showed no comparable improvement. In this context, the large between-group effect size can be viewed as a plausible reflection of the targeted instructional exposure provided to the TG.

Change-score analysis further supported these results. TG participants gained an average of 16 percentage points, while CG participants showed no meaningful change ($M = –0.59$). The between-group difference was statistically significant ($t = 5.53$, $p < .001$, $d = 1.53$). The strength and consistency of these findings, supported by high statistical power (1.00) and strong alignment between content and assessment, increase confidence that participation in EDSTEMA was associated with substantial gains in teacher SCK, while recognizing that the quasi-experimental design limits definitive causal claims.

Taken together, these findings indicate that, within this study, participation in EDSTEMA was associated with powerful and measurable improvements in teachers' SCK, with statistically significant and educationally meaningful gains evident both within the TG and relative to the CG. The consistent improvements across multiple analyses, supported by large effect sizes and high statistical power, suggest that the program likely supported teachers in deepening their conceptual understanding and better preparing them for classroom instruction. These results underscore the potential importance of sustained, content-focused PD in addressing gaps in teacher knowledge and point to the possibility that models like EDSTEMA may help strengthen science education more broadly. However, while content knowledge (CK) is a critical foundation for effective teaching, it does not necessarily guarantee confidence in applying that knowledge in practice. For this reason, the next section examines whether EDSTEMA was also associated with changes in STE, providing insight into how gains in knowledge related to teachers' beliefs about their instructional capabilities.

Science Teaching Self-Efficacy

The findings from this study suggest that participation in EDSTEMA was associated with substantial positive gains in participating teachers' STE. Teachers in the TG showed a statistically significant increase in STE scores from pretest to posttest, with the vast majority demonstrating measurable improvement across both constructs of the instrument. This strong and consistent pattern of growth is consistent with the interpretation that EDSTEMA supported teachers in enhancing their confidence and perceived ability to teach science, even in the face of challenging content or classroom situations. Moreover, the results align with prior research indicating that targeted, content-rich PD is linked to shifts in teacher beliefs, which are closely tied to instructional practices and long-term student outcomes.

The magnitude of this improvement, as reflected in its large effect size, highlights the practical significance of these changes in STE. This pattern suggests that EDSTEMA provided meaningful learning experiences that not only strengthened teachers' content knowledge (CK) but were also associated with increased confidence in their ability to teach science effectively. Given that teacher STE has been consistently linked to stronger instructional quality and higher student achievement in science (Ashton, 1984; Bandura, 1977; Caprara et al., 2006), these findings suggest a potential relevance for student learning outcomes, although this study does not directly test that relationship. In this way, the gains in STE observed among EDSTEMA participants extend beyond individual teacher growth, pointing toward broader implications for classroom practice and the overall improvement of science education.

When comparing the TG to a CG that did not receive the intervention, the results further underscore a positive pattern associated with EDSTEMA participation. While the CG showed little to no change in their STE scores over the same period of time, the TG experienced significantly greater gains. This represents a meaningful difference. Although most participants experienced increases in STE, five teachers exhibited a slight decline (an average of 1.5 points), highlighting the need to investigate individual differences in response to the program.

These variations may stem from contextual factors such as teaching assignment, school-level support, or differences in prior CK. Another possibility is that participation in EDSTEMA created a degree of cognitive disequilibrium for some teachers (Piaget, 1977). As they engaged more deeply with rigorous science content and instructional strategies, some participants may have come to recognize previously unacknowledged gaps in their understanding or skills. This phenomenon aligns with the Dunning–Kruger effect, which suggests that individuals with limited knowledge may initially overestimate their competence, but as their knowledge deepens, they develop a more accurate and more modest self-assessment (Kruger & Dunning, 1999). This pattern may suggest that a temporary decline in self-efficacy does not necessarily reflect a lack of learning, but may be related to a critical shift in self-awareness that often accompanies early phases of cognitive and professional growth.

These findings suggest an important direction for future research: to explore not only which factors predict differential responses to PD, but also how fluctuations in self-perception may reflect deeper, transformative learning processes. Overall, the findings from this study indicate that, within this context, participation in EDSTEMA was associated with significant and practically meaningful gains in STE, supporting the program's potential value in strengthening both confidence and instructional readiness.

The consistent improvements across the TG, coupled with the lack of change in the CG, are consistent with the interpretation that EDSTEMA played an important role in these outcomes, while recognizing that other unmeasured factors may also have contributed. At the same time, variations in teacher responses highlight the complexity of teaching self-efficacy (TSE) as a construct and point to the need for future research on how PD influences individual trajectories of professional growth. Because TSE has been closely linked to student learning, it is important to consider whether the positive changes observed among EDSTEMA participants also corresponded with measurable benefits for their students. The next section addresses this question by examining student academic achievement in science in relation to teachers' participation in EDSTEMA, the final outcome measure of this study.

Student Academic Achievement in Science

The results of this study indicate that students scored significantly higher on the Oklahoma Core Curriculum Test (OCCT) state science assessment in the year after their teachers completed EDSTEMA than in the year before, with average class performance increasing from 70.15% to 79.41%. This nearly ten-point difference represents a large effect size (d = 1.96), suggesting that the observed change in SAA was not only statistically significant but also educationally meaningful. Despite a relatively small sample size, the study achieved full statistical power, reinforcing the robustness of this finding. Importantly, these higher scores were observed even though different cohorts of students were tested before and after the intervention, indicating that the pattern is consistent with the possibility that changes in teacher practice played a role. However, because the cohorts differed and other factors were not controlled, these results should not be interpreted as definitive evidence that the improvements can be attributed solely to participation in EDSTEMA or to specific changes in teacher practice.

These findings support the broader idea that classrooms in which teachers have deeper CK, more effective instructional strategies, and increased confidence are often associated with stronger student outcomes. This is consistent with Hanushek's (1992) assertion that the difference between having an effective versus an ineffective teacher can translate into as much as a full grade-level difference in SAA.

The large effect observed here is aligned with prior work emphasizing the important role teacher quality plays in shaping student outcomes and suggests that programs like EDSTEMA may represent one promising approach for supporting science achievement. Moreover, the alignment between teacher gains in SCK and STE and corresponding improvements in SAA highlights the interconnected nature of teacher learning and student success, pointing to the potential value of investing in comprehensive, long-term professional development.

Situating the Study Within Existing Literature

Taken together, the findings across SCK, STE, and SAA in science suggest a coherent pattern of outcomes within the context of EDSTEMA: teachers demonstrated gains in content knowledge and confidence in their instructional abilities, and their students, in turn, exhibited higher levels of science achievement.

The significant improvements observed among TG participants suggest that targeted, content-rich PD may play an important role in deepening teachers' conceptual understanding. These results take on added significance when considered alongside a broader body of research emphasizing that teacher quality is a key school-based factor associated with SAA (Bolyard & Moyer-Packenham, 2008; Boyd et al., 2008; Darling-Hammond, 2000; Ferguson, 1991).

Teacher Content Knowledge

One noteworthy study that underscores the importance of teacher quality is the Tennessee Student–Teacher Achievement Ratio class size experiment (Project STAR), which examined elementary student performance across dozens of schools and found that the specific teacher to whom a student was assigned had a stronger association with achievement than other factors, such as the school the student attended (Nye et al., 2004). Although the STAR study powerfully demonstrated that teacher quality was more strongly related to achievement than school-level factors, it could not identify the specific characteristics that make a teacher effective.

This is an area in which the present study seeks to contribute to the literature. By showing that a well-designed professional development (PD) program like Every Day STEM for All

(EDSTEMA) was associated with significant gains in teachers' science content understanding (SCK), these findings help to clarify that deep disciplinary understanding may be one key attribute underlying effective teaching. Strengthening teacher content knowledge (CK), as this study suggests, appears possible through PD and may be consequential for students, as indicated by the corresponding pattern of higher student academic achievement (SAA) observed after teachers participated in EDSTEMA. In this way, the present research complements the STAR study by offering evidence that one adaptable teacher attribute—content knowledge—can be supported through PD and may be related to improvements in SAA. These results further support calls for sustained, subject-specific PD as one promising strategy for supporting teacher effectiveness and advancing educational equity in science instruction.

Several other studies have aimed to clarify the teacher characteristics that are associated with SAA. Akiba et al. (2007), using TIMSS data from over four dozen countries, found that SAA tended to be highest when teachers were fully certified, had at least three years of experience, and held degrees in the content area they were teaching. Fetler (1999) and Clotfelter et al. (2007) echoed these findings, identifying teacher experience, certification, and exam performance as key variables that were significantly related to SAA. Similarly, Monk (1994) found that both the quantity of coursework and training in pedagogy were positively associated with student performance on national science assessments. Smaller-scale studies also support these conclusions; for instance, Magnusson et al. (1992) reported a strong association between teachers' SCK and students' conceptual understanding of heat and temperature.

Additional research has consistently shown that teacher certification, which is often used as a proxy for CK and pedagogical content knowledge (PCK), is strongly associated with differences in student learning. Hawk et al. (1985) and Darling-Hammond et al. (2005) both found that students taught by certified teachers significantly outperformed peers taught by uncertified teachers, particularly in math and reading. These findings collectively underscore the critical need to strengthen teacher CK and pedagogical skills.

EDSTEMA addresses this need directly by enhancing teachers' SCK through sustained, targeted PD. The results of this study extend prior research by showing that increases in teacher SCK associated with participation in high-quality PD occurred alongside measurable improvements in student outcomes, thereby contributing to the literature on how PD may relate to classroom impact and student achievement in science.

Teacher Self-Efficacy

While strengthening teacher science content understanding (SCK) has been identified as an essential component of effective instruction, knowledge alone does not fully account for differences in teacher impact on student learning. Equally important are teachers' beliefs in their own abilities, or science teaching self-efficacy (STE), which are closely related to how confidently and persistently they apply their knowledge in the classroom.

Research has consistently shown that teachers with higher levels of STE are more likely to implement innovative strategies, address student misconceptions, and foster engaging learning environments—each of which is often associated with improved student outcomes. Building on the evidence of Every Day STEM for All (EDSTEMA)'s association with gains in SCK, the next section examines how the program was related to teachers' STE, providing further insight into how professional development (PD) experiences may connect to both instructional practice and student achievement.

The findings of this study contribute meaningfully to the growing body of research suggesting that teacher SCK, STE, and student academic achievement (SAA) are closely related.

The observed increase in STE among EDSTEMA participants aligns with and supports the findings of Sandholtz and Ringstaff (2014), who reported that teachers with deeper SCK tend to exhibit greater confidence in their instructional ability and are more likely to design and implement effective lessons. This study not only adds support for that relationship, but also suggests that strategically designed PD programs like EDSTEMA may play a role in strengthening both SCK and STE by providing targeted content and pedagogical learning opportunities.

These results also align with the work of deLaat and Watters (1995), Lakshmanan et al. (2011), Swanson (2010), Tinnin (2000), and Tschannen-Moran and Hoy (2001), who found that higher confidence in content knowledge (CK) is associated with stronger STE, which, in turn, is linked to SAA. The EDSTEMA program was explicitly designed to integrate CK development with pedagogical application, and the significant gains observed in both SCK and STE are consistent with this design. This study adds to the literature by offering evidence from a quasi-experimental design in an elementary context—an area where science confidence and preparation are often limited—and thereby extends the applicability of previous findings, many of which were based on middle- or secondary-level settings.

Moreover, this study addresses a critical gap identified by prior research; while the association between STE and SAA is well documented (Caprara et al., 2006; Sanders & Rivers, 1996; Wright et al., 1997), fewer studies have simultaneously measured changes in both SCK and STE as outcomes of a single intervention. By capturing pre-to-post changes in both areas and examining them in relation to a specific PD program, this study contributes new insights into how STE may be supported and shaped within content-rich PD experiences. The pattern of growth observed among the vast majority of treatment group (TG) participants offers practical, though not definitive, evidence that such programs may be broadly beneficial across a diverse teacher population and points to a promising avenue for future investigation.

This study also aligns with the work of Guskey (2002) and Tschannen-Moran and Hoy (2001), who emphasized that teaching self-efficacy (TSE) is influenced by multiple sources, including PD experiences, and that structured, supportive PD is often associated with higher confidence and stronger instructional competence. The few teachers who did not show gains in STE in this study underscore the point that individual responses to PD are complex and shaped by contextual factors, such as teaching assignment or prior knowledge, which should be considered in future refinements of PD design and implementation.

Linking Science Content Knowledge, Teaching Self-Efficacy, and Student Outcomes

Taken together, these findings highlight the dynamic interplay between teachers' science content understanding (SCK) and their science teaching self-efficacy (STE), both of which are widely viewed as important for shaping classroom practice. By showing that SCK and STE improved concurrently for Every Day STEM for All (EDSTEMA) participants, this study provides evidence that the two constructs do not operate in isolation but are closely related and may reinforce one another in ways that have potential implications for student outcomes.

Recent longitudinal work also points to a more complex, reciprocal relationship between teachers' self-efficacy and their classroom practice. Holzberger et al. (2020) found positive concurrent associations between teachers' self-efficacy beliefs and multiple facets of instructional quality, and their analyses suggested that aspects of instructional quality also predicted subsequent levels of self-efficacy. In other words, self-efficacy may function not only as a potential influence on how teachers teach but also as a consequence of their ongoing experiences in the classroom. This perspective supports interpreting the STE gains observed in

EDSTEMA as both a possible contributor to, and a reflection of, teachers' engagement with more ambitious science instruction.

Understanding this interconnected relationship is essential for explaining how professional development (PD) may translate into measurable gains in student achievement. The findings of this study align with and contribute to the growing body of research that highlights the interrelated nature of SCK, STE, and student academic achievement (SAA). Although these three constructs are distinct, they are closely related and tend to move together within the classroom. Together, they can be understood as a synergistic system in which changes in teacher SCK and STE are associated with shifts in instructional quality and student learning gains.

First, teachers' SCK serves as a foundational element of instructional quality. When teachers possess a deep and accurate understanding of science concepts, they are better equipped to present material clearly, anticipate student misconceptions, and facilitate meaningful inquiry-based learning experiences. Research has consistently shown that strong SCK is associated with more effective instructional strategies and greater confidence in teaching science (e.g., deLaat & Watters, 1995; Lakshmanan et al., 2011; Sanders & Rivers, 1996; Swanson, 2010; Tinnin, 2000; Wright et al., 1997). In the context of EDSTEMA, this was evident in the significant increase in posttest SCK scores, suggesting that the PD effectively strengthened teachers' conceptual understanding.

This growth in SCK appears to coincide with, and possibly reinforce, gains in STE. Teachers who feel knowledgeable in their subject are more likely to believe in their ability to teach it effectively. Bandura's (1977) social cognitive theory posits that mastery experiences, such as improving content knowledge (CK), are among the strongest sources of self-efficacy beliefs. The results from this study are consistent with that perspective, as teachers who participated in EDSTEMA not only improved their SCK but also reported significantly higher STE scores following the program. In this sense, EDSTEMA can be understood as a PD model intentionally aligned with established frameworks regarding efficacy formation, using content-focused mastery experiences, modeled practice, and structured feedback to cultivate teachers' STE. This suggests that content-focused PD can play a role in supporting teachers' confidence in their science teaching capabilities.

Students taught by EDSTEMA participants in the year(s) following their involvement in the program (Year 1) demonstrated higher achievement on the Oklahoma Core Curriculum Test (OCCT) state science assessment than students those same teachers had taught prior to EDSTEMA (Year 0). Because these results are based on different cohorts of students taught by the same teachers at different points in time, the present study cannot determine whether improvements in teacher SCK and STE directly caused the observed gains in student performance. However, the pattern of findings is consistent with prior research indicating that teachers with higher SCK and stronger STE are more likely to employ effective instructional strategies, persist through challenges, and foster more engaging learning environments (Shulman, 1987, 1992; Tschannen-Moran & Hoy, 2001).

In the context of EDSTEMA, increases in SCK and STE occurred alongside improvements in student outcomes. While the design of this study does not allow strong causal claims about how changes in SCK and STE produce higher SAA, the pattern is consistent with a pathway toward positive change in which content-focused PD supports teacher knowledge and confidence, which are, in turn, related to stronger student achievement.

How EDSTEMA's Design Contributed to the Results

Overall, this analysis reinforces the potential value of sustained, well-designed PD programs in supporting both teacher effectiveness and SAA. In the context of science education, where many elementary and middle school teachers report lower STE and SCK, programs like EDSTEMA offer a promising model for scalable, meaningful professional learning (PL) that is associated with tangible differences in student performance.

The significant gains observed in this study suggest that when teachers are provided with structured opportunities to deepen their SCK and build confidence in their instructional abilities, results can extend beyond teachers themselves to their students. These findings emphasize the potential importance of continued investment in long-term, content-rich PD as one strategy for strengthening science education and addressing achievement gaps.

The pattern of improvements in SCK, STE, and SAA observed in EDSTEMA is consistent with the program's intentional design, which aligned with research-based features of effective PD. Decades of research have suggested that well-designed PD has the potential to enhance teacher CK, instructional practice, and student learning outcomes (Borko, 2004; Buczynski & Hanson, 2010; Carpenter et al., 1989; Desimone et al., 2002; Fields et al., 2012; Heller et al., 2012; Supovitz et al., 2000; Yoon et al., 2007). Reflecting these insights, EDSTEMA was developed using key design principles identified in Desimone's (2009) framework, which emphasizes five core elements—content focus, active learning, coherence, sustained duration, and collective participation—as important for meaningful teacher learning and instructional improvement.

EDSTEMA's design reflects these best practices. It offered sustained engagement over time, emphasized active learning around science content, and encouraged collaboration among participants. The program was coherent with participants' curriculum and instructional goals and provided opportunities for teachers to reflect on and apply their learning. These structural features were likely important contributors to the significant gains observed in teacher SCK and STE. Furthermore, the positive shifts in teacher CK and confidence have important implications for student learning, given research consistently showing that teacher quality, characterized by deep CK and strong instructional teaching self-efficacy (TSE), is closely associated with SAA. In this sense, EDSTEMA's thoughtful, research-aligned design offers a plausible explanation for the pattern of outcomes observed in this study and suggests how strategic PD may function as a promising lever for educational improvement.

The sections that follow explore each of the five elements of effective PD identified by Desimone (2009)—content focus, active learning, coherence, sustained duration, and collective participation—and discuss how their inclusion in the EDSTEMA program design was related to teacher and student outcomes. By examining EDSTEMA through this framework, it becomes possible to understand not only what outcomes were observed, but also how the program's design features may have supported them. Each element of Desimone's model provides a lens for analyzing the mechanisms that support teacher learning and the ways that learning can be translated into classroom practice. This analysis situates EDSTEMA within a broader body of research on effective PD, highlighting the program's strengths and pointing to considerations for future program design and implementation.

Content

The significant gains in teacher science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) observed in this study suggest that content-focused professional development (PD) may play an important role in supporting teacher and student learning. Every Day STEM for All (EDSTEMA) was deliberately designed to prioritize science content while also integrating pedagogical strategies, aligning closely with literature that emphasizes the dual importance of content knowledge (CK) and pedagogical content knowledge (PCK) (Shulman, 1987; Carpenter et al., 1989; Fennema et al., 1996; Hiebert et al., 1996; Rhine, 1998). By embedding content-rich experiences within a supportive professional learning community (PLC), EDSTEMA provided teachers with opportunities not only to strengthen their conceptual understanding but also to consider how that knowledge could be translated into classroom practice. In this way, the findings from this study are consistent with arguments that content-focused PD can be an important lever for supporting teacher effectiveness and, potentially, student outcomes in science.

Consistent with Doyle's (1990) theory that teacher knowledge, pedagogy, and student thinking are interconnected, EDSTEMA focused not only on deepening teacher CK but also on developing instructional strategies that teachers could directly apply in their classrooms. This integrated approach resonates with the findings of Garet et al. (1999, 2001), who cautioned that PD focused solely on content without attention to pedagogical skills may not improve—and may actually hinder—teacher practice. In contrast, their research found that PD designed to enhance both CK and instructional skills was associated with strong, positive effects on teacher outcomes.

EDSTEMA's structure provided sustained, content-rich experiences grounded in teachers' instructional context, consistent with Banilower et al.'s (2007) findings that STEM-focused PD balancing CK and pedagogy is linked to improvements in teaching and learning. The results of this study are also in line with research by Yoon et al. (2007) and Guskey and Yoon (2009), who reported that PD programs incorporating CK and PCK were associated with improvements in teacher practice and SAA. The Knowledge Quartet model (Rowland et al.,

2005) offers additional support for the EDSTEMA design, emphasizing that effective teaching depends not only on knowing content, but also on knowing how to teach it in ways that promote deep student understanding—a foundational aim of EDSTEMA's pedagogical components.

At the same time, the literature reminds us that even well-designed content-focused PD must navigate challenges such as limited school resources, cultural barriers, and teacher resistance (Kennedy, 1998; Darling-Hammond et al., 2017). EDSTEMA sought to mitigate these challenges by aligning PD content with teachers' instructional needs, fostering collaboration, and providing sustained follow-up support—strategies advocated by Desimone (2011) and Garet et al. (1999) as important for relevance and long-term sustainability. By incorporating these design features, the program increased the likelihood that teachers would find the learning experiences relevant and be positioned to integrate new ideas into their daily classroom practices in meaningful ways.

Overall, the EDSTEMA program illustrates the potential of PD that integrates strong content knowledge with pedagogical growth, within a supportive and instructionally relevant design. The structure and focus of EDSTEMA reflect many of the key characteristics identified in the literature as important for effective PD, including sustained engagement, collaborative learning, and practical classroom application. The positive results of this study are consistent with decades of research suggesting that content-focused PD, when thoughtfully designed and grounded in teachers' instructional contexts, may contribute to meaningful, lasting improvements in teacher effectiveness and student learning.

Active Learning

The findings of this study support a substantial body of research highlighting the important role of active learning in effective PD. Every Day STEM for All's (EDSTEMA's) structure engaged teachers in applying content knowledge (CK) and pedagogical strategies through hands-on activities, lesson design, and real-world classroom application. These practices are aligned with the constructivist principles of Inhelder and Piaget (1958) and Vygotsky's (1934, 1978) theories of learning. Rather than passively receiving information, participants engaged in reflective practices and collaborative problem-solving that were associated with gains in their science content understanding (SCK), science teaching self-efficacy (STE), and students' academic achievement (SAA). This pattern aligns closely with Garet et al. (1999, 2001), who emphasized the importance of professional development (PD) programs that provide teachers with opportunities to do hands-on work and directly apply learning in their classrooms.

EDSTEMA's focus on real-world implementation allowed teachers to refine their skills through iterative cycles of experimentation and feedback, much like the experiential learning cycle outlined by Kolb (1984). Teachers were encouraged to apply what they learned and were given structured opportunities to reflect on the outcomes, supporting the development of instructional strategies that were both practical and contextually relevant.

Darling-Hammond (1995) argued that PD should engage teachers as both learners and instructors, and EDSTEMA did so by creating an environment in which teachers navigated the dual roles of exploring new content while applying it to their teaching. This experience deepened their pedagogical understanding and exemplifies the principles of Vygotsky's zone of proximal development, in which learning is most effective when support is provided as individuals work just beyond their current capabilities. The growth in teacher confidence observed in this study is also consistent with Bandura's (1977) theory of self-efficacy, which posits that experiencing success from one's actions tends to strengthen belief in one's abilities.

Guskey (1986) further suggested that changes in teacher attitudes and beliefs are more likely to occur after teachers see the successful implementation of new strategies. This perspective highlights the importance of PD designs that move beyond theory to provide opportunities for real-world application and reflection.

The positive shifts observed in this study suggest that EDSTEMA's design, which required teachers to apply and reflect on their learning in the classroom, may have supported the development of a sustained sense of competence and motivation. Penuel et al. (2007) and Yoon et al. (2007) emphasized links between active engagement in PD and improvements in student outcomes, and the SAA results in this study are consistent with those patterns, lending support to the potential value of hands-on, practice-based learning experiences.

Finally, EDSTEMA's inclusion of collaborative elements, such as peer discussions and shared reflection, supported the development of a community of practice (Wenger, 1999), reinforcing the value of collective inquiry and mutual support. This environment appeared to deepen professional learning and sustain teacher engagement over time—an element of effective PD identified by Lieberman (1995) and supported by research on professional development communities (PLCs) and peer learning (Darling-Hammond, 1995; Penuel et al., 2007). By fostering a culture of shared responsibility and ongoing dialogue, EDSTEMA helped ensure that teachers were not isolated in their growth but were supported by a network of colleagues working toward common goals.

Overall, the EDSTEMA experience illustrates how active learning within PD can be associated with changes in teacher knowledge, confidence, and student learning. By enabling participants to learn through doing, reflecting, and collaborating, the program created a sustained, immersive learning environment in which new content and strategies were closely connected to teachers' own classroom contexts, enhancing the perceived relevance and likely retention of what they learned. These findings contribute to the growing body of evidence advocating for active, hands-on PD that is not only reflective but also deeply embedded in the realities of classroom teaching, thereby increasing its potential for long-term impact.

Coherence

The results of this study are consistent with the literature that emphasizes coherence as an important element of effective professional development (PD). Every Day STEM for All (EDSTEMA) was deliberately designed to align with district and school goals, state science standards, and the instructional priorities of participating teachers. This alignment helped ensure that teachers could clearly see the connection between what they were learning in PD and what they were expected to implement in their classrooms, increasing the perceived relevance and applicability of the program. By building upon teachers' prior knowledge while addressing their ongoing instructional needs, the program created a consistent and unified learning experience. The program's coherence occurred alongside measurable gains in science content understanding (SCK) and science teaching self-efficacy (STE) and improvements in student academic achievement (SAA), patterns that are aligned with the broader body of literature on effective PD.

As Desimone (2011) emphasized, coherence helps PD remain relevant, contextually grounded, and applicable to teachers' everyday work. EDSTEMA incorporated these principles by structuring PD sessions around both district expectations and individual teacher goals, which appeared to support teacher engagement and may have increased the likelihood of sustained implementation. This intentional alignment meant that teachers were not simply learning abstract strategies, but rather engaging with practices that directly related to their instructional responsibilities and student needs. Garet et al. (1999, 2001) and Grant et al. (1996) similarly found that PD that is coherent and aligned with curriculum and instructional goals is more likely to be associated with changes in teacher practice, a conclusion that resonates with this study's observed improvements in classroom teaching and instructional confidence.

The cohort-based structure of EDSTEMA also appeared to contribute to its overall coherence. Lieberman (1995) and Leithwood et al. (2004) highlight that collaborative, long-term PD experiences can foster coherence by encouraging teachers to work together over time in pursuit of shared learning goals. In EDSTEMA, this structure allowed participants to revisit ideas, build on shared experiences, and collectively problem-solve challenges, which seemed to support the integration of new knowledge into practice. This sustained, collective engagement helped participants recognize the relevance of new strategies and connect them to broader educational objectives, potentially increasing both teacher investment and classroom effectiveness.

At the same time, the challenges of achieving coherence cannot be overlooked. Hargreaves (2014) cautioned that many PD initiatives fail due to fragmentation and a lack of meaningful alignment with teachers' real-world classroom experiences, which often leads to limited impact. EDSTEMA sought to address this issue by ensuring that new learning was consistently connected to daily instructional practice and reinforced through structured follow-up, peer collaboration, and opportunities for reflection. Teachers were not left to make these connections independently; instead, the program intentionally built in mechanisms for alignment and collective problem-solving that appeared to strengthen coherence, sustain teacher engagement, and reduce the risk of PD becoming an isolated or irrelevant event.

Furthermore, research by Penuel et al. (2007) and Yoon et al. (2007) has emphasized that coherence in PD is often associated with stronger teacher learning and improved student outcomes. The improvements in SAA observed among students of EDSTEMA participants are consistent with this perspective and suggest that PD aligned with curriculum goals, teacher needs, and classroom realities may be more likely to relate to positive student results. These patterns point to coherence as more than a structural feature of effective PD; they suggest it may serve as an important condition that links professional growth to broader educational outcomes.

Overall, the EDSTEMA PD program illustrates the potential role of coherence in supporting teacher growth and student achievement. By ensuring alignment across goals, content, and context, EDSTEMA created a unified professional learning experience that teachers could readily connect to their instructional practice. These findings contribute to the growing body of research advocating for PD that is not only coherent but also sustained, aligned, and strategically integrated into the larger fabric of teaching and learning. Building on this foundation of coherence, the next section examines how the program's sustained duration may have further supported its effectiveness and contributed to lasting improvements in teacher and student outcomes.

Sustained Duration

The outcomes of the Every Day STEM for All (EDSTEMA) program, which included significant improvements in teacher science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA), are consistent with a substantial body of research highlighting the importance of professional development (PD) duration. EDSTEMA's 86-hour commitment, delivered through an intensive summer institute (SI) and follow-up workshops, meets and exceeds recommendations in the literature regarding the amount of time often associated with meaningful professional growth. As emphasized by Yoon et al. (2007), PD programs with fewer than 14 hours showed no statistically significant impact on

SAA, whereas those exceeding this threshold tended to be linked to positive outcomes—especially when supplemented with ongoing activities.

The findings from this study align with Guskey and Yoon's (2009) analysis, which concluded that many of the most effective PD programs included more than 30 hours of sustained, content-rich instruction. EDSTEMA's structure falls within the 50–80 hour plus range identified by Corcoran et al. (2003) as being associated with the strongest outcomes and is consistent with Desimone's (2011) recommendation for a minimum of 20 contact hours.

Moreover, the results are compatible with the argument by Garet et al. (1999, 2001) that extended PD provides teachers with greater opportunities to engage deeply with new content, apply it in their classrooms, reflect, and refine their strategies—practices that were intentionally embedded in the EDSTEMA model. This iterative, long-term approach is further supported by the work of Loucks-Horsley et al. (2003), Weiss et al. (1998), and Penuel et al. (2007), who emphasize the need for sufficient duration to support collaborative learning and sustained instructional change.

The opportunity for extended engagement also echoes findings by Hiebert (1996), Lieberman (1995), Hauck (2012), and Scott et al. (2014), all of whom highlight the importance of long-term PD for fostering instructional improvement. Smith and Neale's (1989) recommendation for multi-year PD to achieve deep change is also relevant; while EDSTEMA was a single-year program, its use of a consistent cohort and follow-up sessions created conditions that may support longer-term development.

At the same time, as Kennedy (1998) rightly cautions, duration alone does not guarantee effectiveness. EDSTEMA's positive outcomes were likely related not only to its sustained duration but also to its intentional design, which paired extended contact time with high-quality, content-focused instruction and active teacher engagement. This perspective underscores the importance of PD that is both well-structured and responsive to teachers' needs, rather than simply longer in length.

Taken together, the EDSTEMA model illustrates a comprehensive and balanced approach that integrates adequate duration with engaging delivery, practical relevance, and collaborative learning—an approach that is supported by decades of research and is consistent with the positive patterns observed in this study.

Collective Participation

The positive gains in science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) from this study are consistent with the idea that collective participation may have played an important role in the Every Day STEM for All (EDSTEMA) experience. This structure fostered ongoing collaboration and peer support, consistent with findings from Ball (1996) and Knapp (1997), who advocate for professional development (PD) models that bring together teachers from similar grades or subjects to promote shared discourse and reflection.

Research by Garet et al. (1999, 2001) reinforces this view, showing that PD featuring collective participation is associated with enhanced teacher practices by enabling sustained, substantive conversations around pedagogy. Within EDSTEMA, participants reported learning not only from the program content but also from one another, which they perceived as contributing to more meaningful and lasting improvements in their instruction.

The collaborative nature of EDSTEMA also appeared to help bridge the gap between theory and practice, echoing Desimone's (2011) and Penuel et al.'s (2007) findings that teachers are more likely to implement strategies when they can discuss and refine them with colleagues.

This shared application of learning encouraged a sense of accountability, a key factor identified by Guskey (2002) as important for sustaining the benefits of PD over time. Moreover, the collegial atmosphere cultivated through the program likely supported teachers' sense of professional satisfaction and motivation.

These outcomes are consistent with Hourd's (1997) work on the long-term value of peer networks. While challenges such as time and scheduling can sometimes hinder collective participation (Hourd, 1997), the EDSTEMA model illustrates how programs that deliberately keep groups intact across sessions may foster benefits such as greater cohesion, mutual support, and extended professional dialogue that can offset these limitations.

Overall, the collaborative design of EDSTEMA aligns with key principles from the research literature and is consistent with the observed gains in teacher growth and student achievement. By engaging teachers in shared learning, reflection, and problem-solving, the program fostered a sense of professional community that extended beyond the PD sessions themselves. This structure appeared to support individual growth while also creating collective momentum, as participants learned from one another and worked to sustain changes in practice over time.

The evidence from this study is in line with a growing consensus that collective participation is an important feature of effective PD design and may be a valuable component of future initiatives aimed at improving both instructional quality and student outcomes. Table 13 summarizes how each EDSTEMA design feature was implemented, the study evidence associated with that feature, and the theoretical mechanisms by which impacts are hypothesized to occur. Collectively, these features not only reflect best practices in PD but also align with core sources of efficacy formation by offering teachers opportunities for mastery experiences, vicarious learning, social persuasion, and a supportive context for interpreting their successes and challenges.

Table 13

How EDSTEMA's Design Contributed to the Results

PD Feature	What EDSTEMA Implemented
Content	86-hr PD emphasizing core science concepts aligned to state standards; misconceptions work using item analyses; content plus pedagogy integration
Active Learning	Hands-on investigations, lesson design; analysis of student work; classroom implementation with feedback
Coherence	Alignment to district goals/standards; direct connection to participants' curricula; follow-up sessions to maintain momentum
Sustained Duration	Intensive summer institute plus structured follow-ups across the year; spaced opportunities to apply and revisit content
Collective Participation	Cohort-based model; peer discussion, shared planning, and reflection; informal community of practice

Note. PD = professional development

Alignment With Established Research on Effectiveness of Professional Development

The design of the Every Day STEM for All (EDSTEMA) professional development (PD) program incorporated several research-based features: extended duration, active learning, a content focus, coherence with standards and goals, and collective participation. Together, these elements were associated with significant improvements in science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). The program's sustained, cohort-based model allowed teachers to engage deeply with science content while actively applying new strategies through collaborative, hands-on experiences. This design aligns with Desimone's (2009) framework for effective PD and is consistent with findings from Garet et al. (2001), Penuel et al. (2007), and Yoon et al. (2007), who emphasize the importance of coherent, content-rich PD that fosters active teacher engagement. By integrating these best practices, EDSTEMA offers an example of how a well-structured PD model can be associated with lasting improvements in teaching and learning, in line with decades of literature that link high-quality PD to both teacher development and student success.

The improvements in SAA observed in this study also align with prior research suggesting that increases in teacher content knowledge (CK) and instructional confidence are often linked to improved student outcomes (Yoon et al., 2007). These findings are consistent with the idea that when teachers are better equipped with strong conceptual understanding and confidence in their teaching abilities, their students are more likely to experience academically beneficial instruction and engaging learning environments.

The improvements in SAA observed in this study also align with prior research suggesting that increases in teacher content knowledge (CK) and instructional confidence are often linked to improved student outcomes (Yoon et al., 2007). These findings are consistent

with the idea that when teachers are better equipped with strong conceptual understanding and confidence in their teaching abilities, their students are more likely to experience academically beneficial instruction and engaging learning environments.

Moreover, the observed effect sizes provide quantitative support for a positive association between aligned, immersive PD models and classroom practice and learning, offering evidence that extends beyond teacher perceptions to measurable student achievement. Taken together, these results reinforce the interconnected nature of teacher knowledge, self-efficacy, and student success, and highlight the potential of sustained, high-quality PD initiatives such as EDSTEMA to support positive change.

In addition, the study aligns with theoretical models of teacher change that emphasize the interconnected development of knowledge, beliefs, and practices (Guskey, 2002; Bandura, 1977). The simultaneous growth observed in both teacher confidence and student performance is consistent with the view that these qualities do not develop in isolation, but rather function in a reciprocal relationship in which changes in one domain are related to changes in the others.

These findings suggest that when teachers are supported in deepening their content understanding and developing effective pedagogical skills, they may be better positioned to foster meaningful student learning that extends beyond surface-level gains. In this way, the results of the EDSTEMA program are compatible with theoretical frameworks that view teacher growth and student achievement as mutually reinforcing processes within the broader system of education.

Overall, this study is consistent with existing research on the potential effectiveness of sustained, content-focused PD in science education. The significant gains observed in teacher CK, teaching self-efficacy (TSE), and SAA align with prior findings that link high-quality PD to improvements in both teacher capacity and student outcomes. These results support established theoretical models of teacher change and underscore the promise of integrated PD approaches that emphasize content mastery, active learning, and classroom application.

Contribution to the Literature: Expanding the Understanding of Effective Science Professional Development

Taken as a whole, this study provides strong empirical evidence that participation in Every Day STEM for All (EDSTEMA) was associated with substantial gains in teacher learning and student achievement, within a sustained, research-aligned professional development (PD) framework that integrated science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA). By documenting sizeable, relatively rapid changes in each of these areas, the findings go beyond traditional expectations of modest PD impacts and point to the potential of well-structured interventions in elementary science education to support meaningful shifts in both teaching and learning.

The results align with prior research on the importance of teacher quality while extending the literature by illustrating how a targeted, content-rich, and coherent PD program can be linked to concurrent enhancements in teacher knowledge, beliefs, and practice alongside higher student performance. In doing so, this study helps bridge a critical gap between theoretical models of teacher change and practical classroom outcomes, offering suggestive evidence that sustained, deeply aligned PD may contribute to meaningful improvements in both teaching and learning. Collectively, these findings contribute to and expand the current understanding of effective PD in elementary science education, positioning EDSTEMA as a promising model for scalable, evidence-informed professional learning.

Exceeding Expectations for Professional Development Outcomes

The literature on professional development (PD) frequently reports modest to moderate gains in teacher and student outcomes, particularly in studies that lack strong implementation fidelity or are constrained by time and context (Yoon et al., 2007; Kennedy, 1998). These tempered expectations have shaped assumptions about the impact PD can realistically have, especially in elementary science, which historically receives less attention and investment than subjects like math or English language arts (ELA).

In contrast, this study documented substantial effect sizes across key variables (SCK, STE, SAA) within a single school year. The treatment group (TG) significantly outperformed the comparison group (CG), which was matched on baseline characteristics, suggesting that this PD model may be highly promising. While the quasi-experimental design does not permit definitive causal conclusions, the pattern of results challenges the notion that only incremental progress is possible through PD and supports the idea that—when well structured—PD can be associated with relatively rapid and meaningful improvements. This expands the field’s understanding of what may be possible when PD design prioritizes coherence, depth, and responsiveness to both teacher and student needs.

Providing Direct Evidence for the Teacher Knowledge to Student Achievement Pathway

Another important contribution of this study is that it provides empirical evidence relevant to the often-theorized but less frequently examined pathway linking teacher content knowledge (CK) and student academic achievement (SAA). While many studies infer that gains in teacher CK may indirectly support student learning, relatively few have measured these constructs together within a single research design.

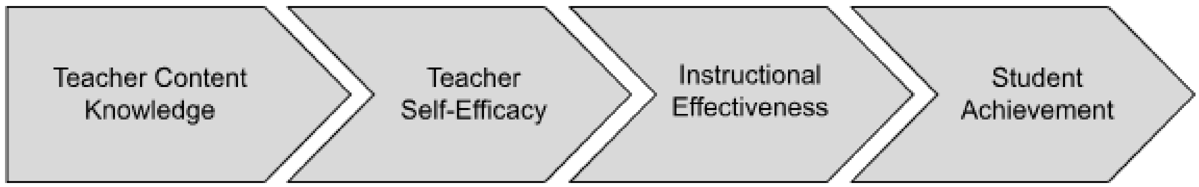
In this study, all three stages—teacher CK, instructional confidence (as TSE), and student performance—were measured, and each showed statistically significant gains associated with participation in Every Day STEM for All (EDSTEMA). In a field where this relationship is commonly assumed but seldom investigated comprehensively, these findings add weight to the rationale for professional development (PD) investments that target deep, sustained improvements in teacher expertise. The results suggest that enhancing teachers' understanding and confidence is not only valuable in its own right but may also be related to stronger student learning outcomes.

The Professional Development Transformation Cascade Model

To conceptualize the interconnected pathway through which professional development (PD) influences student learning, this study introduces the *Professional Development Transformation Cascade Model* (see Figure 4). This model hypothesizes a sequential process in which increases in teacher content knowledge (CK) are linked to improvements in teaching self-efficacy (TSE), which are then associated with enhanced instructional effectiveness and, potentially, improved student academic achievement (SAA). As illustrated in Figure 4, each stage of the cascade builds upon the previous one, emphasizing a potential pathway through which comprehensive PD can support positive change in classrooms.

Figure 4

The Professional Development Transformation Cascade Model



Note. This model depicts a conceptual pathway suggesting how gains in teacher content knowledge (CK) may contribute to increased self-efficacy (TSE), enhanced instructional effectiveness, and, ultimately, higher student achievement (SAA). It is intended as a framework for interpreting the present findings and for guiding future research, rather than as a definitive test of causal relations among these constructs.

Addressing Gaps in Elementary Science Professional Development Research

One of the persistent gaps in professional development (PD) research is the underrepresentation of elementary science, despite national initiatives calling for stronger STEM education starting in the early grades. Many existing studies prioritize secondary instruction or mathematics and literacy, leaving elementary science educators without clear, research-based guidance. This has left elementary educators with limited models for how science-specific PD may influence classroom instruction and student academic achievement (SAA).

This study contributes to addressing that gap by presenting empirical evidence from a science-focused PD program tailored specifically to elementary teachers. By focusing exclusively on science content and pedagogy for elementary educators, the study offers insights into the potential of elementary science PD to support both teacher and student growth. The results highlight the value of prioritizing elementary science in future research agendas and PD funding initiatives.

Moreover, the findings are consistent with the view that elementary educators, when supported with targeted and content-rich PD, can achieve outcomes on par with, or in some cases exceeding, those observed at other grade levels or in other disciplines. This supports the importance of designing PD experiences that are responsive to the specific challenges and needs of early science instruction. In doing so, the study not only addresses a persistent gap in the literature but also illustrates what high-impact PD in elementary science can look like.

Advancing Integrated Professional Development Models

Professional development (PD) programs are frequently fragmented, addressing content knowledge, pedagogical strategies, or classroom context in isolation. Every Day STEM for All (EDSTEMA), however, was intentionally designed to integrate all three components (SCK, PCK, and contextual application) into a cohesive, sustained learning experience. This study provides evidence that such an integrative approach is both feasible and associated with positive outcomes. The gains observed in teacher content knowledge (CK), confidence, and student performance are consistent with the theoretical premise that integrated PD models may be related to stronger and more comprehensive outcomes. These findings suggest that future PD initiatives might benefit from moving away from siloed approaches and instead considering models that intentionally blend the cognitive, practical, and contextual aspects of teaching.

Strengthening Methodological Rigor in Professional Development Research

Another significant contribution of this study lies in its methodological design. Much of the existing professional development (PD) research relies on simple pre–post designs without comparison groups, which limits the ability to distinguish program-related changes from other external influences. By employing a quasi-experimental design with a well-matched comparison group, this study strengthens the credibility of inferences about patterns associated with participation in the Every Day STEM for All (EDSTEMA) program.

This approach enhances confidence in the findings and illustrates that more rigorous experimental logic can be applied within authentic school contexts, where random assignment is often impractical. In doing so, this research provides a useful model for future PD studies, showing how careful design choices can increase methodological rigor while still accommodating the realities of educational practice.

Highlighting the Role of Teacher Self-Efficacy

Teacher self-efficacy (TSE) is a well-established predictor of classroom practice and student academic achievement (SAA), yet it is often omitted from professional development (PD) studies or treated as a secondary outcome. This study measured changes in science teaching self-efficacy (STE) alongside science content understanding (SCK) and SAA, providing a fuller picture of how PD participation was associated with teacher development. The increase in TSE suggests that confidence is not merely a byproduct of knowledge acquisition but may function as an important component of instructional change.

These findings support the inclusion of efficacy measures in PD research and underscore the importance of addressing both cognitive and affective dimensions of teacher growth. The results contribute to the literature by showing that changes in TSE can occur concurrently with knowledge gains and may be related to differences in student outcomes.

Demonstrating the Potential for Rapid Gains

Finally, this study challenges the assumption that substantial gains in teacher practice and student learning can only be realized after multiple years of sustained professional development (PD). Within a single academic year, teachers who participated in Every Day STEM for All (EDSTEMA) exhibited significant improvements in their science content understanding (SCK) and science teaching self-efficacy (STE), and their students showed measurable improvements in student academic achievement (SAA). While the quasi-experimental design does not permit definitive causal conclusions, this pattern of results is consistent with the idea that, under favorable conditions, PD may be associated with relatively rapid and meaningful improvements. It suggests that when programs are well designed, sustained over a reasonable period, and closely aligned with actual classroom needs, substantial progress can be observed within comparatively short timeframes. This insight has practical implications for how PD is planned, funded, and evaluated in school systems seeking efficient, high-leverage interventions.

Implications for Practice

Taken together, these contributions suggest that Every Day STEM for All (EDSTEMA) not only informs theoretical understandings of effective professional development (PD) but also offers actionable insights into how such programs can be designed and implemented in real-world school contexts. By documenting substantial gains in teacher knowledge, confidence, and student achievement within a single year, this study illustrates the potential for PD to serve as one lever for systemic improvement in science education. This section builds on these insights by considering the practical implications of the findings for schools, districts, and policymakers working to address persistent challenges in science teaching and learning.

The findings of this study offer several implications for educational practice, particularly in light of systemic challenges facing science education in the U.S. Issues such as reduced instructional time, inequitable access to high-quality resources, and growing national concerns over scientific literacy highlight the urgent need for targeted interventions. In many parts of the country, science is marginalized relative to mathematics and ELA, with elementary grades often devoting little to no instructional time to science. This early neglect can create knowledge gaps that compound over time, leaving students underprepared for the rigor of secondary science courses.

Furthermore, disparities in school funding translate into stark inequities in resources, curriculum materials, and teacher PD opportunities. These inequities are particularly pronounced in high-poverty schools, where students are less likely to encounter inquiry-based, hands-on science experiences that cultivate curiosity and critical thinking.

National surveys also point to troubling patterns: U.S. students often underperform compared to international peers on measures of scientific literacy, and there is persistent underrepresentation of historically marginalized groups in advanced science coursework and STEM career pathways. These disparities are compounded by systemic inequities in funding, access to high-quality instruction, and opportunities for meaningful engagement in science during the elementary years.

Taken together, these realities underscore that science education in the United States is under significant strain from systemic constraints that undermine both equity and excellence. Without deliberate investment in teacher capacity, more equitable resource distribution, and policies that prioritize science as essential to democratic participation and workforce preparation, the nation risks widening opportunity gaps, reducing global competitiveness, and failing to cultivate scientifically literate citizens capable of addressing complex societal challenges.

The results of this study are consistent with the need for strategic, well-designed PD that not only supports stronger teacher practice but also aligns with broader educational and societal goals. When thoughtfully structured, PD has the potential to help bridge the gap between research and practice by equipping teachers with both the knowledge and confidence to deliver high-quality science instruction. By addressing classroom-level needs while also responding to systemic inequities, PD can serve as one important lever for advancing equity and excellence in science education.

Ultimately, investments in programs like EDSTEMA illustrate how targeted professional learning can be associated with immediate improvements in teacher and student outcomes and can contribute to the longer-term goal of cultivating a scientifically literate citizenry prepared to meet future challenges.

Design Professional Development That Directly Connects to Student Outcomes

The study provides evidence consistent with a pathway linking teacher content knowledge (CK) and teaching self-efficacy (TSE) to student academic achievement (SAA), reinforcing the importance of aligning professional development (PD) efforts with both instructional and learning goals. Rather than treating teacher growth as an isolated objective, PD programs can be intentionally structured to support the translation of knowledge gains into improved classroom practice and, potentially, stronger student performance. Incorporating assessment tools that measure both teacher and student outcomes can help ensure that PD is not only meaningful but also accountable in terms of its real-world effects. By situating SAA as a central consideration in PD design, developers and policymakers can create professional learning experiences that are more likely to be associated with measurable improvements in both teaching quality and student achievement, thereby contributing to a stronger system of science education.

Prioritize Elementary Science in Professional Development Offerings

Given the underrepresentation of elementary science in existing research and professional development (PD), this study points to an important need to invest in science-specific professional learning (PL) for elementary educators. Early exposure to formal science experiences is widely regarded as important for fostering scientific literacy, curiosity, and problem-solving skills, yet elementary teachers often receive less support in this area than in subjects like math and ELA.

Without targeted PD offerings, many teachers may feel they lack the confidence or content knowledge (CK) needed to deliver high-quality science instruction. School districts and education leaders may therefore wish to prioritize the development of PD programs that are specifically designed to meet the needs of elementary teachers. Such efforts have the potential to help address longstanding gaps in instructional quality and to increase the likelihood that students receive consistent and meaningful exposure to science throughout their K–12 education.

Implement Integrated Professional Development Models

The positive results associated with Every Day STEM for All (EDSTEMA) suggest that integrated professional development (PD) models that combine content knowledge (CK), teaching strategies, and real classroom application may be particularly promising. When these components are addressed in a coordinated rather than isolated way, teachers may be more likely to use what they learn in ways that are relevant to their everyday practice. Integrated PD can help connect what teachers know with how they teach and what their students need, making new learning more coherent and easier to apply. Moving away from siloed approaches has the potential to yield more useful and meaningful PD experiences that are better positioned to translate into changes in classroom practice. Going forward, PD programs may benefit from intentionally reflecting the real-world complexity of teaching so that what teachers learn is not merely theoretical, but can be applied to support student learning.

Include Self-Efficacy as a Professional Development Target

Enhancing teacher confidence in their instructional abilities is not just a byproduct of content learning; it is an important outcome in its own right. Teachers who believe in their capacity to teach effectively are more likely to try new strategies, persist through instructional challenges, and engage students more meaningfully. Professional development (PD) programs may therefore benefit from incorporating intentional strategies to build self-efficacy alongside content and pedagogical knowledge. This might include opportunities for practice, feedback, peer collaboration, and reflection—all of which can help teachers internalize their growth and see themselves as capable professionals. When PD supports both what teachers know and how confident they feel applying it, classroom instruction is more likely to be empowered and responsive, with the potential to be associated with improved student outcomes.

Use Rigorous Evaluation Designs

The study's quasi-experimental design adds credibility to its findings by allowing for stronger inferences about patterns associated with the professional development (PD) program than would be possible with simple pre–post designs. Rigorous evaluation methods, such as including comparison groups, help reduce (though not eliminate) the influence of other variables and provide a clearer picture of how outcomes are related to participation in an intervention. PD designers should strive to incorporate such research-informed designs, along with valid and reliable assessment tools that measure both teacher and student outcomes. Collecting and analyzing data systematically enables program developers to identify areas for improvement and make data-informed decisions. Moreover, adopting a rigorous approach to evaluation can increase stakeholder confidence in a program's promise and contribute valuable evidence to the broader field of education research.

Limitations

Despite the valuable insights this study offers, several limitations must be acknowledged when interpreting the findings. These limitations include issues related to participant selection, assessment administration procedures, differential attrition between groups, and issues related to instrument design and scoring. Each of these factors has the potential to influence the results and should be taken into consideration. While the design of this study included thoughtful strategies to avoid or reduce limitations, several notable limitations persist. Acknowledging these issues is essential for understanding the boundaries of the study’s conclusions and for informing the design of future professional development (PD) research that aims to be both effective and generalizable.

Self-Selection Bias

One significant limitation of the study is the presence of self-selection bias, which stems from the lack of random assignment for both the treatment group (TG) and comparison group (CG). Participation in Every Day STEM for All (EDSTEMA) was entirely voluntary, as was participation in the CG. This introduces the possibility that individuals who chose to engage in the study may differ in important ways from those who did not. These differences may include higher levels of intrinsic motivation, a stronger commitment to professional growth, or a greater willingness to devote personal time and effort to improving their science teaching practices.

This self-selection is considered a limitation because it raises concerns about the generalizability of the findings. Specifically, if the participating teachers were more professionally engaged or had fewer external constraints (e.g., personal obligations, school-level barriers), then the pattern of outcomes associated with the program might not be replicable in a broader, more representative sample of elementary teachers. In other words, the magnitude of gains in science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement (SAA) may have been influenced not only by participation in the professional development (PD) but also by the characteristics of the teachers who chose to opt in.

As noted by Shadish et al. (2002), self-selection can pose a threat to validity—particularly when individuals who participate in an intervention differ systematically from those who do not—thereby complicating both the interpretation and generalization of results.

In this study, self-selection bias may have resulted in a TG composed disproportionately of teachers who were already more confident in their capacity to benefit from science-focused PD or who had more robust support systems (e.g., collaborative school cultures, flexible administrators, or access to instructional resources). Such individuals may have also been more inclined to persist through the challenges of an 86-hr PD program, complete an extensive application process, and implement new instructional strategies. Similarly, the CG was not randomly selected but rather recruited voluntarily and compensated per assessment, which may have attracted a distinct subset of teachers, including those motivated by incentives rather than a sustained commitment to science teaching improvement.

Mitigating the Impact of Self-Selection Bias

To mitigate the impact of self-selection bias, several steps were taken in the design and analysis of the study. First, efforts were made to ensure that the comparison group (CG) had similar characteristics to the treatment group (TG). Both groups were compared at baseline on key variables such as prior science content understanding (SCK) scores, years of teaching experience, and demographic factors. Statistical analyses confirmed no significant differences between the groups on these measures, thereby supporting the comparability of the groups at the beginning of the study. This strengthens the plausibility that the observed posttest differences can be attributed to participation in Every Day STEM for All (EDSTEMA) rather than to preexisting differences between participants.

Nonetheless, the voluntary nature of participation remains a limitation that should be acknowledged when interpreting results and considering the scalability of the program. Future researchers could address this limitation by employing random assignment wherever feasible, such as using a lottery system when demand for professional development (PD) exceeds available slots. Alternatively, researchers could implement *stratified sampling* or *propensity score matching* to construct CGs that more closely reflect the broader population of elementary teachers (Shadish et al., 2002). Collecting qualitative data on participants' motivations for enrolling could also help identify and account for potential differences in engagement or mindset. Finally, replication of this study in a variety of school settings, including those with less voluntary participation, would help test the robustness and transferability of EDSTEMA's effects.

School District Classification

Another limitation concerns district type. Every Day STEM for All (EDSTEMA) was implemented in a consortium of suburban districts with relatively similar organizational structures and proximity to university partners. Implementation in rural districts might be complicated by travel distance, lower staffing levels, multi-grade teaching assignments, and fewer opportunities for in-person collaboration. In contrast, urban districts may need to adapt the model to accommodate larger cohorts, higher mobility, and layers of district- and school-level initiatives. Future research should investigate how professional development (PD) models like EDSTEMA function in rural and urban settings and identify design features that support successful adaptation across diverse contexts.

Attrition

A significant limitation of this study involves participant attrition, particularly within the comparison group (CG). Of the 44 teachers originally recruited for the CG, only 26 completed both the pretest and posttest for the science content understanding (SCK) assessment, and just 17 completed both administrations of the Science Teaching Efficacy Beliefs Inventory Form A (STEBI-A). This level of attrition is noteworthy, as it resulted in a 61% reduction in usable data for the STEBI-A measure and a substantial decrease for the SCK assessment as well. Such missing data restricts the size of the sample, limits the generalizability of the results, and raises concerns about potential bias in the final data set. These issues compromise the reliability of findings and decrease statistical power, making it more difficult to detect true effects of the intervention.

Attrition is a limitation because it introduces the risk that the final sample of participants does not represent the broader population originally targeted. If the teachers who dropped out of the study differed in meaningful ways from those who remained—such as being less motivated, having more demanding schedules, or being less confident in their science instruction—the findings could be skewed toward more favorable outcomes. This is especially problematic for internal validity, as any observed differences between the treatment group (TG) and CG may reflect preexisting group differences rather than the true effect of the Every Day STEM for All (EDSTEMA) professional development (PD) program.

Mitigating the Impact of Attrition

To mitigate the impact of attrition, a pairwise deletion strategy was used to minimize the loss of usable data without significantly interfering with other variables in the study. The method was used under the assumption that the missing data were missing at random (MAR) (Little &

Rubin, 1987, as cited in Shadish et al., 2002), and while this assumption cannot be fully verified, it allowed for a more inclusive and consistent approach to analysis, albeit with some risk of residual bias. Still, the magnitude of missing data represents a serious limitation. It reduces the comparability between groups and undermines the strength of any inferences drawn from the results. Furthermore, smaller and inconsistent sample sizes limit the external validity of the findings, as they may not be generalizable to the broader population of elementary science teachers, especially those less likely to fully participate in professional development (PD) initiatives or research studies.

Future researchers should consider employing strategies to reduce attrition and to account for its potential effects on validity and generalizability. Practical approaches may include implementing follow-up incentives to encourage sustained participation, offering more flexible assessment windows to accommodate teachers' schedules, or employing intent-to-treat analyses in which all participants initially assigned to a group are included in the analysis regardless of completion status. Collecting more detailed demographic and contextual information on non-completers would also allow researchers to better determine whether missing data are truly MAR or instead *missing not at random* (MNAR)—a distinction that carries significant implications for the strength of inferences drawn from the data. Addressing attrition in these ways would not only strengthen the internal validity of future studies but also enhance confidence that the findings accurately represent the broader teacher population.

Science Content Knowledge Assessment

Another important limitation of this study involves the use of the Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) to measure teacher science content understanding (SCK). Although the MOSART assessments were chosen for their strong reliability, alignment with national science standards, and relevance to the content emphasized in the Every Day STEM for All (EDSTEMA) program, they were originally designed for K–12 student populations, not adult learners. As such, the instruments have not been explicitly validated for use with elementary educators, and this mismatch introduces concerns about the appropriateness and sensitivity of the assessments in accurately capturing teachers' depth of science knowledge and conceptual understanding.

This presents a limitation because adult learners, such as practicing teachers, may possess different prior knowledge, reasoning skills, and misconceptions than children. As a result, MOSART assessments may not fully reflect the complexity or subtlety of teacher understanding, nor adequately measure the type of professional knowledge that supports effective instruction. While the assessments provide a useful proxy for conceptual knowledge and offer insight into persistent science misconceptions, their application to adult participants may result in under- or overestimations of teacher learning. This limitation could impact the interpretation of study results by introducing measurement error, potentially reducing or inflating the observed effects of EDSTEMA on teacher SCK.

Mitigating the Impact of the Science Content Knowledge Assessment

To mitigate this limitation, careful consideration was given during the selection of assessment items. The Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) instruments were chosen because no validated science content understanding (SCK) assessments currently exist for elementary educators, and the content covered by the MOSART closely aligned with both the goals of the Every Day STEM for All (EDSTEMA) professional development (PD) program and the relevant science standards. Additionally, previous research (e.g., Sadler et al., 2013) has shown that misconceptions-based assessments, even when initially designed for students, can be effective in identifying gaps in teacher content knowledge (CK) and pedagogical content knowledge (PCK). These factors supported the decision to use MOSART as the most appropriate available tool for this study. However, the lack of adult validation still remains a notable constraint when interpreting findings.

Future researchers should consider developing and validating SCK assessments specifically tailored to elementary educators. These tools should account for adult learning theory, the unique structure of teacher misconceptions, and the kinds of content knowledge (CK) most relevant for effective instruction at the elementary level. Moreover, triangulating results with additional measures such as interviews, lesson artifacts, or performance-based assessments could provide a more comprehensive picture of teacher learning. Until more specialized instruments are available, researchers using student-oriented tools like MOSART with adults should acknowledge this limitation and interpret results with appropriate caution.

If future researchers elect to use MOSART assessments to measure SCK, they should only select questions from the middle school MOSART subtests and avoid using high school-level questions, based on information gathered from item-level analysis. While most items showed improvement on the posttest, a few declined—particularly those aligned with high school-level standards and which required higher levels of cognitive processing. These results may point to content areas of the EDSTEMA program that require further scaffolding or instructional support. The difficulty of these items and possible guessing behaviors on both administrations suggest that variability in scores may not reflect a regression of conceptual understanding, but rather uncertainty or lack of familiarity with more advanced topics.

Self-Efficacy Inventory

Another limitation of this study involves the use of the Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) to measure changes in teacher self-efficacy. While the STEBI-A is a widely recognized, validated instrument used across various educational research contexts, it is important to acknowledge that it relies on self-reported data, which introduces potential bias. Respondents' answers may be influenced by factors such as social desirability, inaccurate self-perception, or a desire to align responses with perceived program goals. As a result, teachers may either overestimate their science teaching efficacy due to enthusiasm from participating in the professional development (PD) program or underestimate it due to a heightened awareness of their own knowledge gaps, which were revealed during the learning process. This variability can impact the reliability and validity of the results.

Moreover, although the STEBI-A has been validated in numerous educational settings, it has not been specifically tested in the context of sustained, cohort-based PD programs such as Every Day STEM for All (EDSTEMA). Its unique structure, which included features such as extended duration, collaborative learning model, and intensive science content focus, may introduce additional factors not accounted for by the instrument. For example, teachers engaged in a long-term, content-rich PD experience may experience shifts in efficacy beliefs that are unexpected but normal, such as temporary dips in confidence due to confronting misconceptions. These nuanced patterns may not be fully captured by the STEBI-A, which was not designed to assess self-efficacy development over time.

Mitigating the Impacts of the Self-Efficacy Inventory

To mitigate these limitations, the STEBI-A was administered as both a pretest and posttest, allowing for a comparison of relative change in self-efficacy over the course of the program rather than relying on end scores. This approach helped focus the analysis on directional shifts rather than static beliefs, which can reduce the impact of individual bias to some extent. Additionally, findings were interpreted alongside the science content understanding (SCK) data to provide a more triangulated understanding of changes in teacher learning and confidence.

Future researchers could pair self-report measures like the STEBI-A with qualitative or performance-based assessments, such as classroom observations, reflective journals, or structured interviews. These additional data sources could provide deeper insights into how teachers' efficacy beliefs evolve and how those beliefs translate into instructional practice. Furthermore, future studies might consider modifying or extending the STEBI-A to better align with the unique features of cohort-based or content-intensive professional development (PD) models, ensuring that the instrument captures a fuller spectrum of teacher experience and growth.

Pretest and Posttest Administration Procedures

One significant limitation of this study involves the administration procedures of the pretests and posttests, particularly for the comparison group (CG). Unlike the treatment group (TG), which completed assessments in structured and monitored environments, CG participants were provided with assessment links to complete asynchronously under unmonitored conditions. This lack of supervision introduced several threats to the validity and reliability of the data collected from the CG. Unmonitored testing environments carry an inherent risk of assessment contamination. Without supervision, there is no assurance that participants refrained from using external aids such as internet searches, textbooks, or assistance from other individuals. The potential for participants to access unauthorized resources compromises the integrity of the test results and introduces bias, as the assessment may no longer accurately reflect participants' actual science content understanding (SCK) or science teaching self-efficacy (STE). Additionally, participants may have been more focused on receiving the $100 compensation provided for completing the assessments rather than engaging with the content thoughtfully. This extrinsic motivation could have led to disengaged behaviors, such as randomly selecting answers or rapidly completing assessments without genuine effort, which would further reduce the reliability of the data.

These uncontrolled testing conditions are particularly problematic when interpreting between-group comparisons. If CG participants were able to artificially inflate their scores through the use of external resources or other shortcuts, the observed difference between the TG and CG could be underestimated. This would weaken the perceived impact of the Every Day STEM for All (EDSTEMA) program, potentially masking the full extent of its effectiveness on teacher learning outcomes. In contrast, it is also possible that some CG participants may have lacked accountability or intrinsic motivation, and thus may have underperformed due to a lack of engagement, which would have confounded the results. Either scenario introduces error and weakens the internal validity of the study.

Mitigating the Impacts of the Pretest and Posttest Administration Procedures

To mitigate this limitation, the study employed clear instructions emphasizing the importance of honest and independent work, and participants were informed that the assessments were intended to measure their actual knowledge to help inform instructional practices. This framing was designed to encourage intrinsic motivation and to reduce the likelihood of participants seeking outside assistance during the assessments. However, the lack of accountability measures such as proctoring, lockdown browsers, or in-person administration may have limited the effectiveness of these mitigations and left room for potential response biases.

Future research would benefit from implementing more rigorous controls during assessment administration to ensure the integrity of assessment data, thereby strengthening confidence in the validity of the findings. One possible solution is the use of remote proctoring software that monitors screen activity and webcam footage to ensure test integrity. Alternatively, researchers could arrange in-person or synchronous online proctored sessions for all participants to maintain consistent conditions across groups. Moreover, reducing or eliminating financial incentives tied strictly to assessment completion may help ensure more authentic engagement with the assessment tasks. Establishing standardized, controlled testing conditions is essential for enhancing the accuracy and validity of group comparisons in studies of this nature.

Student Academic Achievement Data

A significant limitation of this study involves the use of student academic achievement (SAA) data derived from the Oklahoma Core Curriculum Tests (OCCTs) during a period of transition within the state’s Department of Education. Specifically, the timing of the study coincided with the rollout of a revised version of the science OCCT. At the time of data collection, validity and reliability data for the newly implemented assessment were not yet available. This poses a considerable challenge to interpreting the findings with confidence, as any changes in student performance between years may not solely reflect the influence of teacher participation in Every Day STEM for All (EDSTEMA), but could also be attributed to variations in the assessment instrument itself.

The absence of established psychometric data for the new OCCT version creates uncertainty about the consistency and fairness of the assessment. Differences in test structure, content alignment, item difficulty, or scoring rubrics between the old and new versions could have significantly influenced students’ test performance independent of instructional quality.

This lack of comparability between Year 0 and Year 1 test administrations undermines the internal validity of the study and limits the ability to draw firm conclusions about the relationship between teacher participation in EDSTEMA and subsequent student gains. Without assurance that the assessment instrument measured student knowledge in a stable and consistent manner across years, it is unclear whether observed score changes are the result of actual improvements in student learning or artifacts of the revised testing format.

Additional limitations related to OCCT data include the self-reporting of student scores by treatment group (TG) teachers. Because the comparison group (CG) did not submit corresponding SAA data, cross-group analyses were not possible. This limitation restricts the scope of inference and confines interpretations of student outcomes to TG scores only. Moreover, test scores are subject to annual fluctuations based on a variety of contextual factors, including student cohort characteristics, school policies, and changes in curriculum pacing or instructional time, all of which can influence outcomes independently of teacher quality or professional development (PD) participation.

Another important limitation is the absence of a fourth-grade science OCCT, which meant that SAA data could not be collected for students taught by fourth-grade teachers in the TG. This omission resulted in an incomplete dataset and constrained the ability to assess the full impact of the EDSTEMA PD program across all grade levels represented in the study. As a result, the findings primarily reflect outcomes for fifth-grade teachers and their students, leaving open questions about whether similar gains might have occurred at the fourth-grade level. Future research should address this gap by incorporating alternative measures of student achievement for grades without state science assessments, ensuring a more comprehensive evaluation of PD effectiveness across the elementary spectrum.

Mitigating the Impact of Student Academic Achievement Data

To mitigate these limitations, efforts were made to ensure accurate and honest reporting of student test scores by treatment group (TG) participants, and only scores from official state assessments were included in the analysis. These steps helped to strengthen the reliability of the data by reducing the risk of inflated or fabricated reporting. However, in the absence of independently verified and fully comprehensive datasets, the possibility of error or bias cannot be completely ruled out. For this reason, caution must be exercised when interpreting the student achievement outcomes, and future studies should prioritize obtaining verified records directly from state or district databases to enhance validity.

Future researchers should aim to conduct similar studies after the Department of Education has fully validated and published psychometric data for the revised science Oklahoma Core Curriculum Test (OCCT). Additionally, they should attempt to obtain student assessment data directly from school districts or the state testing authority to ensure accuracy and reduce reliance on self-reported outcomes. Including both treatment and comparison group data, and ensuring uniform assessment conditions across grade levels, would also enhance the robustness of future analyses. Where standardized assessments are unavailable, alternative validated measures of student science achievement, such as performance-based assessments or curriculum-embedded tests, could provide more consistent and reliable indicators of learning gains.

Integration of Professional Development Features

Another important limitation of this study is its inability to determine which of the features of effective professional development (PD) described in the literature had the greatest impact on the observed outcomes. While the Every Day STEM for All (EDSTEMA) program was intentionally designed to incorporate five research-based features of effective PD (content focus, active learning, coherence, sustained duration, and collective participation), the design of the study did not allow for the individual contribution of each feature to be examined in isolation. This limitation arises from the integrated and holistic nature of the PD program itself. All five features were implemented concurrently, and participants engaged with them as part of a unified learning experience. As a result, it is difficult to disentangle the specific influence of each individual feature on the outcomes measured. This lack of differentiation restricts the interpretability of the findings and limits the ability to identify which elements were most responsible for improvements in teacher content knowledge (CK), teaching self-efficacy (TSE), or student learning. Without such insight, practitioners and policymakers seeking to replicate or scale similar PD models may not know which components are essential, which can be adapted, and which may have minimal impact.

The inability to identify the most impactful features also affects the study's practical utility. Although the overall PD model was associated with positive outcomes, the inability to attribute those outcomes to specific PD features hinders efforts to streamline or optimize professional learning experiences for efficiency or context-specific needs. This limitation underscores a common challenge in PD research: while multifaceted programs often yield strong outcomes, their complexity makes it difficult to determine precisely what works and why.

Mitigating the Impact of the Integration of Professional Development Features

To mitigate this limitation, the study clearly documented the presence and implementation of each of the five features in the Every Day STEM for All (EDSTEMA) professional development (PD) model. This documentation provides transparency and allows readers to understand the comprehensive structure of the intervention. However, no methodological strategy was employed to parse out the relative influence of each feature. Future researchers should consider study designs that allow for the disaggregation of professional development (PD) components. For example, employing a factorial or mixed-methods design could help isolate and compare the effects of individual features (Shadish et al., 2002). Alternatively, conducting a series of PD interventions that each emphasize different combinations of features would allow for a comparative analysis of which elements are most critical for improving science content understanding (SCK), science teaching self-efficacy (STE), or student academic achievement (SAA). Incorporating participant reflections, fidelity measures, or observational data may also help illuminate which aspects of PD teachers find most transformative and why. Understanding the relative contributions of each feature would significantly enhance the field’s ability to design more targeted and effective professional learning experiences.

Limitations Key Takeaways

- Self-selection persists despite baseline equivalence; because participation was voluntary, the findings may not generalize beyond this particular group of teachers, even though baseline characteristics were similar across groups.
- Differences in assessment administration likely affected comparison group (CG) scores; as a result, TG–CG comparisons should be interpreted cautiously, with the understanding that some portion of the difference may reflect testing conditions rather than the professional development (PD) alone.
- Attrition in the CG reduced statistical power and may have introduced bias; the analyses assume data are missing at random (MAR), but the possibility of MNAR cannot be ruled out.

- Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART) and Science Teaching Efficacy Beliefs Instrument–Form A (STEBI-A) provided useful, but imperfect, measures for an adult PD context; future work would benefit from triangulating these instruments with qualitative or performance-based data (e.g., observations, teacher reflections, or instructional artifacts) to strengthen confidence in the results.
- The Oklahoma Core Curriculum Test (OCCT) data constrained interpretations of student academic achievement (SAA) outcomes, as scores were available only for the TG, were collected during a state assessment transition and did not include fourth-grade teachers.
- The integrated PD design likely contributed to the positive outcomes observed, but it also limited the ability to determine which specific features were most influential; future studies should consider factorial, component, or mixed-methods designs to better isolate the effects of individual PD elements.

To aid readability, Table 14 summarizes the study's major limitations in the order presented—self-selection, administration procedures, attrition, instrument design and scoring, and the integrated PD model—along with the mitigation steps taken, implications for interpretation, and concrete recommendations for future research.

Table 14

Summary of Study Limitations

Limitation Area	What/Where Occurred	Risk	Implications
Self-Selection Bias	Voluntary TG & CG; no random assignment	TG/CG may differ (motivation, support); threatens internal validity & limits generalizability	May over/ underestimate observed effects
Assessment Administration Procedures	CG completed assessments asynchronously/ unmonitored; TG monitored	Possible use of outside resources or low-effort responding could lead to biased CG scores; weakens TG–CG comparisons	Between-group differences may be under/overestimated
Attrition	CG SCK & STEBI-A had high pretest noncompletion	Reduced power; final sample not representative of original carefully chosen cohort; missing data may not be MAR	Estimates may be biased; external validity reduced
SCK Measure (MOSART)	Instrument designed for K–12 students, not adults	Possible misfit for teacher knowledge; measurement error	SCK gains could be under/overestimated
Self-Efficacy Measure (STEBI-A)	Self-report; reverse-coded items noisier	Social desirability & wording effects could introduce reliability concerns	Direction and magnitude of STE change should be interpreted cautiously
Student Achievement (OCCT)	State test transition; CG OCCT not collected; no grade 4 science test	Psychometrics uncertain; no TG–CG student comparison; incomplete grade coverage	SAA inferences TG-only; year-to-year shifts may reflect test changes
Integration of PD Features	Five features implemented concurrently	Cannot isolate which feature drives effects	Pragmatic design yet restricts clarity on causal mechanisms

Note. TG = treatment group; CG = comparison group; SCK = science content knowledge; STEBI-A = Science Teaching Efficacy Beliefs Form A; MAR = missing at random; MOSART = Misconceptions-Oriented Standards-Based Resources for Teachers; STE = science teaching efficacy; OCCT = Oklahoma Core Curriculum Tests; PD = professional development; SAA = student academic achievement.

Future Directions

Although the current study provides strong evidence that participation in Every Day STEM for All (EDSTEMA) was associated with positive changes in teachers' science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement

(SAA) in science, it also raises important questions for continued exploration. These results point to the promise of sustained, content-focused professional development (PD) while also highlighting the complexity of measuring and maintaining teacher growth and SAA over time. There are several ways in which this line of research could be further developed and refined to deepen understanding of what makes PD effective, particularly in elementary science, where research remains limited. Future directions should focus on strengthening the design, evaluation, and theoretical frameworks of high-quality professional development programs, ensuring that they are both empirically rigorous and practically responsive to the needs of teachers and students.

Explore Long-Term Outcomes

Future research should examine the long-term effects of sustained professional development (PD) participation. As Smith and Neale (1989) observed, meaningful instructional change often requires two to three years of continued engagement. While this study measured short-term outcomes, longitudinal research could determine whether gains in teacher content knowledge (CK) and efficacy persist, grow, or fade over time, and how ongoing supports such as peer collaboration, coaching, and reflective practice contribute to lasting improvements in teaching and student achievement.

Investigate Causes for the Decrease in Self-Efficacy

Additionally, future research should aim to better understand the individual differences in teacher responses to professional development. The slight decrease in science teaching self-efficacy (STE) scores observed among five participants in this study suggests that not all teachers may have benefitted from professional development (PD) in the same way or at the same pace. While these differences may reflect variations in prior experience, instructional beliefs, or the level of support available in their school contexts, they may also indicate a deeper cognitive shift.

In line with the Dunning–Kruger effect, some participants may have initially overestimated their teaching efficacy due to limited exposure to high-quality science instruction. As they engaged more deeply with the content and instructional practices presented in Every Day STEM for All (EDSTEMA), they may have become more aware of the complexity of effective science teaching and, as a result, more critical of their own skills. This temporary dip in self-efficacy could signal a productive stage of growth in which teachers move from unconscious incompetence to conscious awareness of their learning needs.

A constructivist lens further suggests that these declines may reflect moments of cognitive disequilibrium. When teachers encounter ideas, tasks, or models of instruction that conflict with their existing beliefs and routines, their prior schemas for "what works" in science teaching may no longer feel sufficient. In such moments, teachers face the choice of assimilating the new information into existing frameworks or accommodating by revising those frameworks altogether. Accommodation is often accompanied by uncertainty and lowered confidence, even as it represents deeper learning. Qualitative studies that capture teachers' reflections at multiple points during and after PD could illuminate how they experience and resolve this disequilibrium.

Qualitative follow-up studies, such as teacher interviews, reflective journals, or case studies, could provide valuable insight into these shifts and help researchers and PD designers understand how to support teachers through this process. A multiple case study design focusing specifically on teachers whose STE scores declined could examine the interaction among content learning, beliefs about teaching, school context, and perceived support. Longitudinal mixed-methods research that combines repeated self-efficacy measures over time with observational data and narrative accounts would be especially useful in tracing whether temporary declines in STE are followed by subsequent rebounds as teachers gain mastery experiences and integrate new practices into their instruction. Understanding these individual trajectories may inform the development of more responsive, differentiated PD models that better support all participants as they progress along their professional learning paths.

Tease out the Individual Impact of Various Professional Development Features

Another critical direction involves exploring which specific features of professional development (PD) are most impactful on different outcomes. As this study implemented all five research-based features of effective PD (content focus, active learning, coherence, sustained duration, collective participation, and alignment with standards), it was not possible to determine which of these components had the most influence on science content understanding (SCK), science teaching self-efficacy (STE), or student academic achievement (SAA). Future studies could employ experimental or quasi-experimental designs that isolate or vary individual features to better understand their unique contributions. For example, factorial designs or mixed-methods process evaluations could examine how the presence or absence of specific features (e.g., duration vs. active learning) affects participant outcomes.

It is also important to consider how duration and quality interact in PD effectiveness. While longer programs like Every Day STEM for All (EDSTEMA) are often recommended, Kennedy (1998) cautioned that simply increasing the number of contact hours is not always beneficial, particularly if the content is poorly aligned with teachers' needs or delivered in disengaging ways. This reinforces the need for PD models that balance adequate duration with high-quality, relevant content and engaging delivery. Future research could explore how different combinations of duration, content focus, and instructional strategy influence teacher engagement and learning over time.

Moreover, building on the work of Garet et al. (1999, 2001) and Guskey and Yoon (2009), future studies should continue to investigate whether the format or structure of PD (e.g., workshops vs. coaching vs. institutes) matters less than the overall design quality and contextual fit. Rather than seeking a universally "best" format, researchers should focus on identifying the conditions under which specific formats are most effective. This includes examining how PD aligns with classroom realities, the degree to which it promotes teacher collaboration, and whether it supports practical implementation of instructional strategies.

Investigate the Link Between Teacher Science Content Knowledge, Science Teaching Self-Efficacy, and Time Spent Teaching Science

Future research should explore whether improvements in elementary teachers' science content understanding (SCK) and science teaching self-efficacy (STE) result in increased time spent teaching science, and whether this, in turn, contributes to gains in student academic achievement (SAA) and the reduction of educational inequities. While this study provides evidence that participation in high-quality professional development (PD) like Every Day STEM for All (EDSTEMA) is associated with significant improvements in teachers' SCK and STE, it remains unclear whether these gains are accompanied by increases in the amount of instructional time dedicated to science in elementary classrooms. This is a critical area of inquiry, as the declining emphasis on science instruction in elementary education has been well-documented. Research shows that early exposure to science is essential for building foundational knowledge, nurturing curiosity, and developing critical thinking skills. If teachers who feel more confident and better prepared to teach science are more likely to increase science instructional time, this could have a powerful impact on both student engagement and achievement. Importantly, such a shift could also help address long-standing inequities in science education. Students from low-income, minority, or otherwise marginalized backgrounds often have less access to informal science learning opportunities outside of school. Increasing time spent on science instruction—particularly in schools serving disadvantaged populations—may serve as a key lever for closing achievement gaps and promoting greater equity in STEM education.

Address Limitations of Current Study

To enhance the rigor and generalizability of future research, it is essential to address the methodological limitations encountered in the present study. One critical improvement would be the use of randomized assignment where feasible, as this would reduce self-selection bias and strengthen causal inferences about program impact. Future studies should also implement controlled and standardized testing environments to improve the validity of assessment data, minimizing the risk of external influences on test performance. In addition, reducing attrition through sustained engagement strategies—such as offering incentives, building ongoing communication, and providing flexible participation options—would ensure more complete datasets and more reliable analyses. Finally, collecting comprehensive and independently verified student outcome data across multiple grade levels would allow for a fuller evaluation of professional development (PD) impact, offering stronger evidence of how programs like Every Day STEM for All (EDSTEMA) influence learning across the elementary spectrum.

Future Related Studies

While the current study offers valuable evidence of the positive impact of the Every Day STEM for All (EDSTEMA) professional development (PD) program on teacher science content understanding (SCK), science teaching self-efficacy (STE), and student academic achievement

(SAA), it also serves as a launching point for further inquiry. Despite some methodological limitations, this study lays a strong foundation for future research on effective science PD. Building on these findings, future studies can deepen and extend the work by exploring the long-term sustainability of PD outcomes, disentangling which program components are most influential, and investigating how shifts in teacher self-efficacy may reflect deeper cognitive or professional growth. The following proposed studies aim to refine, expand, and strengthen the current research by addressing these unanswered questions through varied designs and methodological approaches.

Longitudinal Study of Sustained Impact of EDSTEMA

One important direction for future research is to conduct a longitudinal study that examines the sustained impact of Every Day STEM for All (EDSTEMA) over multiple years. While the current study documented immediate gains in science content understanding (SCK) and science teaching self-efficacy (STE), it is essential to understand whether these gains persist over time and how they evolve as teachers continue to apply what they learned in their classrooms. A longitudinal design could track participants over three years and collect annual measures of teacher SCK and STE, as well as student academic achievement (SAA) data for each year the teacher remains in the classroom. In addition to quantitative assessments, qualitative data such as teacher interviews, reflective journals, or classroom observations could provide insight into how teachers internalize and implement professional development (PD) content over time. Comparing these outcomes with those of a matched comparison group (CG) of non-participating teachers would allow researchers to assess the enduring effects of EDSTEMA participation. This type of study could also identify which forms of ongoing support (e.g., mentoring, professional learning communities, instructional coaching) are most effective in reinforcing and sustaining the initial gains made through PD. Ultimately, a longitudinal investigation would contribute to a more comprehensive understanding of the lasting value of high-quality PD and inform the design of sustained, scalable models for teacher learning.

Analysis of the Effectiveness of Individual Professional Development Features

Another valuable avenue for future research involves conducting a component analysis to identify which specific features of the Every Day STEM for All (EDSTEMA) professional development (PD) program are most influential in improving teacher and student outcomes (Ward-Horner & Sturmey, 2013). While the current study implemented five research-based features of effective PD—content focus, active learning, coherence, sustained duration, and collective participation—it was not designed to isolate the individual contributions of each feature. As such, it remains unclear which components had the greatest impact on science content understanding (SCK), science teaching self-efficacy (STE), or student academic achievement (SAA). A future study could systematically vary the presence or emphasis of these features across different cohorts of participants. For example, one group could receive PD emphasizing content focus and active learning, while another would focus on collective participation and coherence. By comparing outcomes across groups, researchers could identify which combinations of features lead to the greatest gains. This approach would also allow for the examination of potential interactions between features, offering a more nuanced understanding of how effective PD works. Data collection would include pre- and post-assessments of SCK and STE, student performance data, and participant feedback on the perceived value of different PD elements. The findings of such a study would have important implications for the design and implementation of future PD programs, enabling educators and policymakers to prioritize the most impactful features and allocate resources more strategically. Ultimately, this research could help move the field toward more targeted, efficient, and effective professional learning models.

Case Study of Decreased Self-Efficacy

A third future research direction involves conducting a qualitative case study focused on participants who experienced a decrease in science teaching self-efficacy (STE) after completing Every Day STEM for All (EDSTEMA). While the majority of participants in the current study showed positive gains, a small subset experienced slight declines in STE. Rather than viewing these decreases as indicative of program ineffectiveness, it is important to consider alternative explanations.

One such explanation is the Dunning–Kruger effect, which suggests that as individuals gain knowledge, they may become more aware of the complexity of a subject and more realistic in assessing their own abilities. In this context, participation in EDSTEMA may have created a sense of cognitive disequilibrium for some teachers, prompting them to reassess their confidence in light of newly acquired insights into effective science teaching. A multiple case study design could explore this phenomenon by examining the experiences of teachers whose STE scores declined, using interviews, reflective journals, and post- professional development (PD) follow-up assessments.

This research could provide valuable insight into how shifts in self-efficacy may actually mark a productive stage of professional growth. Understanding how teachers process and respond to these moments of disequilibrium could inform the design of PD programs that offer better emotional and cognitive support during periods of intense learning. In doing so, future PD models could be made more responsive to the complex, nonlinear process of teacher development.

Investigate the Link Between Teacher Science Content Knowledge, Science Teaching Self-Efficacy, and Time Spent Teaching Science

To investigate the relationship between science content understanding (SCK), STE, instructional time, and student academic achievement (SAA), a longitudinal mixed-methods study is recommended. The study would recruit a diverse sample of elementary schools, including those serving historically marginalized populations, and collect data over the course of two academic years. Teachers would first complete pre- and post-assessments measuring SCK and STE, both before and after participating in a high-quality professional development (PD) program focused on science instruction. Classroom observations, lesson plans, and weekly teacher logs would be used to track the amount of instructional time allocated to science throughout the school year.

To assess SAA, standardized science assessments and classroom assessment tasks would be administered at multiple points during the year. The data would allow researchers to examine whether increases in teacher SCK and STE are associated with greater time spent teaching science, and whether that additional time is linked to gains in SAA.

Quantitative analyses would examine the relationships among SCK, STE, instructional time, and student achievement. Qualitative data, such as teacher interviews and classroom observations, would provide insight into how changes in teacher confidence and content knowledge influence instructional decisions and pedagogical practices.

Importantly, the study would also analyze subgroup outcomes to determine whether increases in science instruction help narrow achievement gaps among students from different backgrounds. This comprehensive approach would provide valuable evidence on how teacher development affects not only teaching practice but also student learning and educational equity, offering guidance for policymakers, PD designers, and school leaders committed to improving science education in elementary schools.

Summary of Key Findings

The results showed that teachers who participated in Every Day STEM for All (EDSTEMA) demonstrated significant gains in science content understanding (SCK), with a large effect size. These findings are consistent with prior research suggesting that content-focused professional development (PD) is often associated with enhanced teachers' subject-matter understanding (Desimone, 2009; Kennedy, 1998). Participating teachers also experienced statistically significant increases in science teaching self-efficacy (STE), although a small subset exhibited slight declines. Student academic achievement (SAA) in the treatment group (TG) also differed from that of the comparison group, with posttest scores approximately 20 percentage points higher. The between-group effect size (d = 4.93) was extremely large, indicating a substantial difference in performance between the two groups during the year of the study and suggesting a strong association between participation in EDSTEMA and student outcomes, while not, on its own, establishing a causal effect of the PD.

However, interpretation of these results must consider several important limitations. Comparison group participants completed assessments under unmonitored conditions, which may have compromised data validity. The study also coincided with changes to the state science assessment (OCCT), for which no reliability or validity data were yet available. Additionally, due to the integrated design of EDSTEMA, it was not possible to isolate the unique effects of each PD feature, nor to control fully for contextual variables such as teaching assignment or school-level support. Finally, because participation in EDSTEMA was voluntary, self-selection bias may have contributed to the observed differences.

Conclusions

The findings of this study offer meaningful insight into outcomes associated with the Every Day STEM for All (EDSTEMA) professional development (PD) program, but they must be interpreted within the broader context of the study's design, theoretical framework, and limitations. This section synthesizes the results in light of existing research and conceptual models of teacher learning to draw conclusions about what the findings suggest—and what they do not. In doing so, it situates the evidence within a larger scholarly conversation, clarifying how the results align with, extend, or diverge from prior studies on teacher PD and student achievement. It also considers possible explanations and implications for the observed outcomes for both theory and practice, offering a balanced perspective that acknowledges the strengths and constraints of the study while identifying directions for continued inquiry.

The Professional Development Transformation Cascade Model: Linking Content Knowledge, Self-Efficacy, and Student Achievement

The first three conclusions of this study align with the conceptual pathway illustrated in Figure 4, the Professional Development Transformation Cascade Model. This model depicts how professional development (PD) may support a progressive sequence of change: as teachers deepen their science content understanding (SCK), they may experience increased science teaching self-efficacy (STE), which is associated with stronger instructional effectiveness and, in turn, higher student academic achievement (SAA). The empirical results of this study are consistent with this cascade, indicating that gains in teacher knowledge and confidence tended to occur together and were accompanied by improvements in student learning. The following three conclusions elaborate on each stage of this process, illustrating how content-rich, sustained PD can be related to strengthened teacher capacity and measurable benefits for students within this context.

Conclusion 1: Content-Rich, Coherent Professional Development Programs Can Strengthen Teacher Science Content Knowledge

The large increase in science content understanding (SCK) among treatment group (TG) participants is consistent with findings from prior studies suggesting that professional development (PD) with a strong content focus is particularly effective (Garet et al., 2001). These results offer reasonably strong evidence that Every Day STEM for All (EDSTEMA) was associated with meaningful growth in one of the most critical dimensions of effective science teaching: deep subject-matter understanding.

This finding supports the theoretical premise that content knowledge (CK) is foundational to instructional quality (Shulman, 1986; Ball, 1996). When teachers possess a richer and more accurate understanding of the science content they teach, they are better positioned to anticipate student misconceptions, provide clear explanations, and design learning experiences that promote conceptual understanding rather than rote memorization.

The strength of the SCK gains observed in this study also reinforces the argument that elementary teachers—who often receive limited preparation in science—may benefit substantially from PD that prioritizes subject-matter expertise. Moreover, the consistency of these gains across a diverse group of participants suggests that well-designed, content-focused PD can be broadly beneficial, not just for teachers who already have strong science backgrounds. Taken together, these results provide compelling evidence that content-rich PD can serve as an important lever for improving teacher capacity and may, in turn, be related to stronger student learning.

Conclusion 2: High-Quality Professional Development Can Support Teacher Self-Efficacy, but Not Uniformly

While most teachers in this study experienced increased science teaching self-efficacy (STE), a minority showed slight declines. This pattern suggests that STE is a dynamic construct rather than a fixed trait—one that may temporarily dip as teachers confront and acknowledge their own learning needs. This phenomenon is consistent with the Dunning–Kruger effect (Kruger & Dunning, 1999), which proposes that individuals with limited knowledge may initially overestimate their competence, but as their understanding deepens, they develop a more accurate and often more modest self-assessment.

These results also align with research indicating that early stages of learning can prompt cognitive disequilibrium and temporary reductions in confidence before longer-term growth occurs (Bandura, 1977; Piaget, 1977). In the context of professional development (PD), such shifts may actually be a marker of progress, reflecting increased self-awareness and a willingness to critically evaluate one's practice. Accordingly, PD designers may need to anticipate and support teachers through this developmental process by providing scaffolds, feedback, and opportunities for reflection that help participants reinterpret initial uncertainty as part of productive professional growth and, over time, consolidate that growth into more stable confidence and improved instructional practice.

Conclusion 3: Teacher Knowledge and Efficacy Are Related to Student Learning

The treatment group (TG)'s significantly higher posttest performance is consistent with the theory that teacher quality is a powerful school-based factor associated with student academic achievement (SAA) (Darling-Hammond, 2000; Yoon et al., 2007). In this study, the pattern of results suggests that well-structured professional development (PD) programs, such as

Every Day STEM for All (EDSTEMA), may strengthen teacher capacity in ways that are related to differences in classroom practice and student outcomes.

While caution is warranted due to limitations in testing procedures and the reliance on self-reported data, the consistency of positive findings across science content understanding (SCK), science teaching self-efficacy (STE), and SAA provides suggestive evidence of a coherent pathway in which high-quality PD is associated with better-prepared teachers who, in turn, teach students who achieve at higher levels. Because of self-selection and the presence of uncontrolled variables, this relationship should be interpreted as indicative of a possible mechanism rather than as definitive proof of a causal chain. Future research that incorporates randomized designs and formal mediation analyses will be essential for more clearly identifying and testing the pathways through which PD is related to both teacher and student outcomes.

Additional Findings and Broader Implications for Professional Development Design and Implementation

Together, the first three conclusions support the relationships depicted in the PD Transformation Cascade Model, suggesting that strengthened content knowledge (CK) and STE are associated with measurable differences in SAA. However, the effectiveness of PD also depends on the specific features of its design and the contexts in which teachers implement new practices. The following conclusions and discussion broaden the focus to these additional factors, highlighting implications for PD structure, equity, and sustainability.

Conclusion 4: The Five Features of Effective Professional Development Work Well in Combination, but More Research Is Needed to Isolate Their Individual Effects

Every Day STEM for All (EDSTEMA) incorporated all five core features of effective professional development (PD) identified by Desimone (2009)—content focus, active learning, coherence, sustained duration, and collective participation—but the design of the current study did not make it possible to determine which of these elements were most influential in relation to the observed outcomes. This limitation is important, as prior research has suggested that not all PD features contribute equally to teacher learning or student achievement, and their relative influence may vary depending on context, implementation quality, and participant characteristics (Desimone, 2009; Kennedy, 1998). For instance, while content focus and active learning are often emphasized as central drivers of change, coherence and collective participation may play subtler but equally important roles in supporting long-term growth.

To refine understanding of which features, or combinations of features, are most essential, future research should aim to systematically vary and test individual PD elements. Such studies would provide valuable insights into how to optimize PD design, helping ensure that limited resources are invested in strategies that are most strongly associated with positive teacher and student outcomes across diverse educational contexts.

Conclusion 5: Professional Development Is Not One-Size-Fits-All, and Teacher Outcomes Vary Based on Individual and Contextual Factors

The observed variation in teacher responses to Every Day STEM for All (EDSTEMA) suggests that individual characteristics such as prior content knowledge (CK), instructional beliefs, teaching assignment, and school-level support may play an important role in shaping how professional development (PD) is experienced and what outcomes are realized. This pattern aligns with Borko’s (2004) assertion that teacher learning is inherently context-dependent and influenced by the complex environments in which educators work. Some teachers may thrive when presented with rigorous new content, while others may require additional scaffolding, mentoring, or time to fully integrate new knowledge and practices into their instruction.

These differences underscore the importance of designing PD programs with built-in flexibility and differentiation so they can address the varied needs of teachers at different stages of professional growth. By acknowledging and planning for this variability, PD developers can create more inclusive and responsive models that are better positioned to support positive outcomes across diverse teacher populations.

Broader Implications for Theory, Policy, and Research

From a theoretical standpoint, this study is consistent with frameworks that conceptualize teacher learning as situated, reflective, socially mediated, and grounded in active engagement (Desimone, 2009; Piaget, 1972; Vygotsky, 1978). The results provide concrete examples of how teachers' growth in a sustained, content-rich, and collaborative professional development (PD) context can reflect these theoretical principles in practice. This alignment not only lends additional support to these models but also illustrates their applicability in the context of elementary science education, a domain where teachers often report lower levels of confidence and preparation.

From a policy and practice perspective, the findings suggest that strategic investments in robust, research-aligned PD may be associated with meaningful improvements in instructional quality and student outcomes. At the same time, the study highlights the danger of oversimplification: policymakers should be cautious about generic, one-size-fits-all mandates and instead prioritize nuanced, context-sensitive efforts that account for local needs, teacher diversity, and systemic inequities. In particular, programs that integrate science content understanding (SCK) with pedagogical growth may serve as important levers for both strengthening instruction and supporting educational equity.

From a research standpoint, this study underscores the importance of employing long-term, mixed-methods, and contextually rich approaches to better understand how PD operates, for whom it is most effective, and under what conditions it is associated with sustained impact. Such designs can capture the complexity of teacher learning, illuminate variation in individual responses, and reveal the contextual factors that shape implementation and outcomes.

In conclusion, while this study contributes valuable evidence to the fields of professional development and science education, it also points toward the need for more detailed and refined investigations into the mechanisms of teacher growth. The journey toward more effective PD is iterative, requiring continuous cycles of design, implementation, reflection, and improvement. This study represents one step along that path, offering practical insights and theoretical contributions that can inform future efforts to support teacher learning and, in turn, promote student success in science.

Recommendations

Based on the findings of this study, as well as the broader body of relevant literature and the limitations acknowledged within this research, several recommendations are proposed for strengthening professional development (PD), informing educational policy, and guiding future inquiry. These recommendations are grounded in the well-established understanding that teacher quality is among the most influential school-based factors affecting student academic achievement (Darling-Hammond, 2000; Yoon et al., 2007).

As this study illustrates, thoughtfully designed PD may play an important role in supporting teacher knowledge, confidence, and instructional effectiveness, which are in turn often associated with stronger student outcomes. Therefore, the recommendations that follow emphasize not only how PD can be improved in structure and delivery, but also how it can be positioned within larger policy frameworks and future research agendas to support long-term impact.

Recommendation 1: Prioritize Sustainable Funding for Elementary Science Teacher Professional Development

Given the critical role that teacher quality plays in influencing student academic achievement (SAA), policymakers must prioritize sustained investment in professional development (PD) opportunities for educators—particularly for elementary science teachers, who often enter the classroom with limited preparation in this content area. Despite decades of research affirming the importance of highly qualified science educators and the proven impact of targeted PD on both teacher effectiveness and student learning outcomes, federal funding for such initiatives has sharply declined since the discontinuation of the MSP program in 2016. Since the end of MSP, no federally funded program of comparable scale or scope has been implemented to systematically support science teacher development. This absence represents a major policy gap that threatens to stall, or even reverse, progress in improving elementary science instruction at a time when nurturing scientific curiosity and literacy is more critical than ever for both civic participation and workforce readiness.

To address this gap, it is imperative that policymakers reinstate a program similar to MSP or design a new funding stream explicitly dedicated to sustained, content-focused science PD for elementary educators. Such an investment would not only advance national STEM education goals but also ensure that all students—regardless of background—have equitable access to high-quality science instruction from their earliest years of schooling.

Recommendation 2: Invest in Deep Content-Focused Professional Development

The findings of this study reinforce a central truth in education: teachers matter—a great deal. Prior research has consistently shown that teacher quality is the single most influential school-based factor affecting student achievement (Darling-Hammond, 2000), yet defining the specific characteristics that make a teacher “effective” has remained an ongoing challenge.

This study contributes to that conversation by providing evidence that strong science content understanding (SCK) is an important and measurable component of effective teaching, particularly in elementary classrooms where teachers often report limited confidence in science. The results suggest that educational leaders and policymakers may wish to prioritize professional development (PD) programs that immerse teachers in rich, discipline-specific content learning opportunities, rather than relying solely on more generic training.

Such efforts are especially urgent in elementary science, where many teachers lack formal preparation in the subject and where early exposure to high-quality science instruction can profoundly shape students’ curiosity, confidence, and long-term achievement in STEM. By investing in targeted, content-rich PD, policymakers can directly strengthen teacher effectiveness and, in turn, ensure that more students receive the foundational science education necessary for success in later schooling and beyond.

Recommendation 3: Design Professional Development That Is Sustained, Coherent, and Aligned to Teachers' Needs

The success of Every Day STEM for All (EDSTEMA) adds to the growing body of research that underscores the value of professional development (PD) that is sustained over time, grounded in authentic classroom practice, and meaningfully aligned with teachers’ instructional contexts (Desimone, 2009; Garet et al., 2001). In contrast, short-term, one-off workshops—

though common—rarely produce lasting changes in teacher knowledge, beliefs, or practice. Instead, school districts and state education agencies should prioritize PD models that are deliberately structured to unfold across multiple sessions throughout the school year, incorporate collaborative learning structures, and provide opportunities for teachers to actively apply and reflect on their learning. Such models create the conditions for iterative growth, allowing teachers to refine instructional strategies over time while receiving ongoing support.

At the same time, it is important to recognize that longer PD alone is not automatically more effective; duration must be paired with high-quality design, coherence with curriculum and standards, and contextual relevance to teachers' real-world challenges (Kennedy, 1998; Guskey & Yoon, 2009). When these elements come together, PD can move beyond compliance-driven experiences to become a powerful lever for meaningful instructional improvement and student achievement.

Recommendation 4: Monitor and Support Teachers Through Periods of Cognitive Disequilibrium

The observation that a small subset of teachers experienced declines in science teaching self-efficacy (STE) underscores that professional growth is not always linear or immediately affirming. As teachers engaged with new content and reflected critically on their instructional practices, some likely became more aware of previously unrecognized gaps in their knowledge and skills.

This pattern aligns with both Piaget's (1977) concept of cognitive disequilibrium—where confronting challenging ideas can temporarily unsettle confidence—and the Dunning–Kruger effect (Kruger & Dunning, 1999), which suggests that initial exposure to expert knowledge may prompt individuals to recalibrate their self-assessments downward as they develop a more accurate understanding of a domain. Far from signaling failure, these temporary dips in efficacy may represent necessary milestones in the growth process, marking the transition from surface-level confidence to deeper professional competence.

At the same time, cognitive disequilibrium is not uniformly beneficial. When teachers encounter challenging content or unfamiliar instructional approaches without sufficient guidance, time, or encouragement, uncertainty can become discouraging rather than productive, potentially leading to withdrawal from the professional development (PD) or resistance to instructional change. PD designers and school leaders therefore have a responsibility to anticipate that some teachers will experience periods of lowered self-efficacy and to treat these moments as predictable features of ambitious learning rather than as problems to be avoided.

The presence of cognitive disequilibrium does not, in itself, guarantee positive outcomes. Disequilibrium can be productive when it is accompanied by adequate support, time, and resources to work through new ideas and experiment with new practices. In such cases, temporary uncertainty can motivate inquiry, reflection, and refinement of instruction. Conversely, when teachers encounter challenging content or unfamiliar pedagogies without sufficient guidance or encouragement, disequilibrium may be experienced as threatening or overwhelming. Under these conditions, lowered self-efficacy may lead to withdrawal from the PD experience or resistance to change, rather than to growth.

To support teachers through this developmental stage, PD should incorporate structured opportunities for reflection, peer discussion, and targeted coaching that help participants make sense of their evolving beliefs and practice. Tools such as reflective journals, collaborative inquiry groups, and follow-up classroom-based coaching can normalize feelings of uncertainty, highlight evidence of growth that may not be immediately visible in self-ratings, and scaffold teachers as they try new strategies in their classrooms. By intentionally monitoring and supporting teachers during periods of cognitive disequilibrium, districts and PD providers can help ensure that temporary dips in confidence become catalysts for sustained improvement rather than barriers to change.

Recommendation 5: Balance Standardization With Flexibility in Professional Development Design

While the Every Day STEM for All (EDSTEMA) model adhered to the five widely recognized features of effective professional development (PD), the study's findings also suggest that individual teacher responses vary in important ways. Teachers inevitably bring diverse levels of prior content knowledge, beliefs about teaching and learning, and contextual factors shaped by their school environments. As Borko (2004) and Desimone (2009) argue, teacher learning is highly context-dependent, meaning that a one-size-fits-all approach to PD may not fully capture or sustain growth for all participants.

Therefore, PD should retain its core design principles while incorporating flexibility that allows for differentiated pathways and responsive supports. Future PD initiatives might include differentiated tracks tailored to teacher experience, optional extensions for those seeking more advanced engagement, or embedded assessments to identify needs in real time and adjust programming accordingly. Such adaptations would help ensure that PD programs are both equitable and effective, meeting teachers where they are while still holding all participants to high expectations for growth.

Recommendation 6: Use Mixed-Methods and Longitudinal Designs in Future Research

To address the limitations of the current study, future research should adopt longitudinal, mixed-methods approaches that capture both the durability and depth of professional development (PD) outcomes. Tracking teachers and their students over multiple years would provide valuable insight into the sustainability of PD effects and whether initial gains in teacher knowledge, self-efficacy, and student achievement persist over time. Complementing quantitative measures with qualitative data such as interviews, classroom observations, and teacher reflections would help uncover the mechanisms of change and provide a richer understanding of how PD translates into instructional practice.

In particular, factorial or component-based studies could disentangle the relative contributions of different PD features, allowing researchers to determine which design elements are most impactful, for whom, and under what conditions. Such approaches would strengthen both the theoretical grounding and the practical applicability of PD research, ultimately leading to more targeted and effective models for teacher learning.

Recommendations Summary

These recommendations are intended to guide ongoing efforts to strengthen teacher professional development, enhance student learning outcomes, and deepen our collective understanding of what makes professional development (PD) effective. By building on the findings of this study while acknowledging its methodological and contextual constraints, educators, policymakers, and researchers can continue to refine approaches that both elevate teacher practice and promote equitable student achievement in science and beyond. In particular, investing in sustained, content-rich, and contextually responsive PD has the potential to address persistent challenges in science education while also advancing broader goals of educational equity and excellence. Ultimately, the lessons drawn from this study can inform the design of future PD models that are not only effective in the short term, but also sustainable in cultivating teacher expertise and fostering long-term student success.

Concluding Remarks

The current state of science education in the United States presents urgent challenges, many of which are rooted in long-standing systemic issues such as reduced instructional time for science, inequitable access to high-quality learning experiences, and persistent gaps in scientific literacy among both students and the general public. As scientific understanding becomes increasingly vital for informed citizenship and personal decision-making, the nation's lagging performance on international science assessments and widespread science illiteracy are causes for concern. These issues are particularly pressing at the elementary level, where early exposure to science is critical for fostering curiosity, critical thinking, and future interest in STEM careers. However, pressures from standardized testing, limited instructional time, and a lack of teacher confidence have all contributed to the marginalization of science in the elementary curriculum.

This study was designed in response to those challenges. It explored the impact of the Every Day STEM for All (EDSTEMA) professional development (PD) program—an initiative specifically designed to address gaps in elementary science teaching by strengthening teacher science content understanding (SCK), enhancing science teaching self-efficacy (STE), and ultimately improving student academic achievement (SAA). Drawing on the five widely recognized features of effective PD (Desimone, 2009), EDSTEMA provided teachers with sustained, content-rich, and collaborative learning experiences aligned with research-based best practices. The results of this study offer robust evidence that this kind of approach is associated with significant improvements in both teacher and student outcomes.

Importantly, this research is consistent with the view that teacher quality remains one of the most critical school-based factors related to student success, echoing decades of educational research (e.g., Darling-Hammond, 2000). While previous studies have struggled to identify exactly what constitutes an “effective” teacher, this study adds to a growing body of literature suggesting that deep content knowledge is a foundational component of teacher effectiveness, particularly in science. By supporting teachers in strengthening their understanding of core scientific concepts and their confidence in teaching them, EDSTEMA appeared to help narrow the gap between teacher preparation and classroom practice—an essential consideration for improving science education outcomes at scale.

Moreover, this study addresses an important gap in the literature: the limited research specifically evaluating the effectiveness of elementary science PD programs. While many PD initiatives have aimed to improve general teaching practices, relatively few have focused explicitly on science, and fewer still have examined links between teacher PD and both teacher-level and student-level outcomes using a quasi-experimental design. This study not only documents meaningful gains in teacher SCK and STE, but also identifies substantial differences in student science achievement associated with teachers’ participation in EDSTEMA, providing a strong rationale for continued investment in thoughtfully designed science-focused PD.

At the same time, the study acknowledges the complexity of teacher learning. Not all participants responded uniformly to the PD experience—some experienced temporary declines in efficacy, potentially due to increased awareness of their learning gaps, a phenomenon aligned with the Dunning–Kruger effect. These nuances serve as a reminder that teacher growth is nonlinear and context-dependent, and that PD programs must be flexible and responsive to the varied needs of participants.

In sum, this study contributes important new knowledge to the field of science education by providing evidence that carefully structured PD is associated with positive changes in both teacher learning and student outcomes, even within the constraints of today’s educational system.

While limitations such as self-selection bias, unmonitored assessments, and evolving state testing systems caution against overgeneralization, the results are nonetheless promising. They suggest that addressing the root causes of underperformance in science—such as limited teacher preparation, lack of confidence, and inequitable access to high-quality instruction—begins with equipping teachers with the tools, knowledge, and support they need to succeed.

By responding directly to the challenges outlined in Chapter 1, this study reinforces the idea that meaningful improvement in science education is possible. Effective professional development is not a silver bullet, but when thoughtfully designed and implemented, it can serve as a powerful lever for change. As the nation continues to grapple with questions of educational equity, scientific literacy, and workforce readiness, ensuring that all students have access to excellent science instruction, and that all teachers are prepared to provide it, remains a critical priority.

References

ACT. (2008). *The forgotten middle: Ensuring that all students are on target for college and career readiness before high school*. https://www.act.org/content/dam/act/unsecured/documents/ForgottenMiddle.pdf

AICBT (2025). *AI–Therapy Statistics* (BETA version) [Comparing two sets of data]. https://www.ai-therapy.com/psychology-statistics/results/20250222181103573

AICBT (2025). *AI–Therapy Statistics* (BETA version) [Comparing two sets of data]. https://www.ai-therapy.com/psychology-statistics/results/20250222185040563

AICBT (2025). *AI–therapy statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222152527919

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222153222088

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222153650088

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222154700205

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222155339468

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222160450443

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222160948727

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222180412355

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222191737246

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222211502134

AICBT (2025). *AI–Therapy Statistics* (BETA version) [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222212217239

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 37*(7), 369–387. https://doi.org/10.3102/0013189X073087

Ashton, P. (1984). Teacher efficacy: A motivational paradigm for effective teacher education. *Journal of Teacher Education, 35*(5), 28–32. https://doi.org/10.1177/002248718403500507

Ball, D. L. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan, 77*(7), 500–508. https://www.jstor.org/stable/i20405614

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1982). Self-efficacy mechanisms in human agency. *American Psychologist, 37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the national science foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching*, *44*(3), 375–395. https://doi.org/10.1002/tea.20145

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc. https://horizon-research.com/NSSME/wp-content/uploads/2020/04/Report_of_the_2018_NSSME.pdf

Barton, P. E., & Coley, R. J. (2008). *Windows on achievement and inequality*. ETS Policy Information Center. https://files.eric.ed.gov/fulltext/ED501477.pdf

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*(6), 830–847. https://doi.org/10.1002/sce.21078

Bleicher, R. E. (2004). Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers. *School Science and Mathematics, 104*: 383–391. https://doi.org/10.1111/j.1949-8594.2004.tb18004.x
.
Bolyard, J. J., & Moyer-Packenham, P. S. (2008). A review of the literature on mathematics and science teacher quality. *Peabody Journal of Education*, *83*(4), 509–535. https://doi.org/10.1080/01619560802414890
.
Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Teacher preparation and student achievement* (NBER Working Paper No. W14314). National Bureau of Economic Research. http://ssrn.com/abstract=1264576

Braun, H., Coley, R., Jia, Y., & Trapani, C. (2009). *Exploring what works in science instruction: A look at the eighth-grade science classroom*. ETS Policy Information Center. https://www.ets.org/Media/Research/pdf/PICSCIENCE.pdf

Brenner, P. S., & DeLamater, J. (2016). Lies, damned lies, and survey self-reports? Identity as a cause of measurement bias. *Social Psychology Quarterly*, *79*(4), 333–354. https://doi.org/10.1177/0190272516628298

Buczynski, S., & Hansen, C. B. (2010). Impact of professional development on teacher practice: Uncovering connections. *Teaching and Teacher Education, 26*(3), 599–607. https://doi.org/10.1016/j.tate.2009.09.006

Capps, D. K., & Crawford, B. A. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, *24*(3), 497–526. https://www.jstor.org/stable/43670581

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44*, 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499–531. http://www.jstor.org/stable/1162862

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement?* (Working Paper 12828). National Bureau of Economic Research. http://www.nber.org/papers/w12828

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Corcoran, T., McVay, S., & Riordan, K. (2003). *Getting it right: The MISE approach to professional development* (CPRE Research Report Series RR-055). Consortium for Policy Research in Education. https://files.eric.ed.gov/fulltext/ED498324.pdf

Council of State Science Supervisors. (April 2006). *Roundtable discussion*. Meeting of the Council of State Science Supervisors. Washington, DC.

Creswell, J. W. (2014). *Research design: Qualitative, Quantitative, and Mixed Method Approaches* (4th ed.). Sage.

Dailey, D. D. (2013). *The effects of a STEM professional development intervention on elementary teachers* (Order No. 3587609) [Doctoral dissertation, University of Arkansas at Little Rock]. ProQuest Dissertations & Theses Global. https://search-proquest-com.ezproxy.lib.ou.edu/docview/1427342722?accountid=12964

Darling-Hammond, L. (1995). Changing conceptions of teaching and teacher development. *Teacher Education Quarterly, 22*(4), 9–26. http://www.jstor.org/stable/23475817

Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan, 78*(3), 193–201. https://www.jstor.org/stable/20405748

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

Darling-Hammond,L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. Teachers College Press.

Darling-Hammond, L., Flook, L., Schachner, A., & Wojcikiewicz, S. (2022). Educator learning to enact the science of learning and development. *Learning Policy Institute*. https://files.eric.ed.gov/fulltext/ED617576.pdf

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*(42), 1–51. https://doi.org/10.14507/epaa.v13n42.2005

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://files.eric.ed.gov/fulltext/ED606743.pdf

deLaat, J., & Watters, J. J. (1995). Science teaching self-efficacy in a primary school: A case study. *Research in Science Education*, *25*(4), 453–464. https://doi-org.ezproxy.lib.ou.edu/10.1007/BF02357387

Denis, D. J. (2016). *Applied univariate, bivariate, and multivariate statistics.* Wiley.

Derrick, B., Toher, D., & White, P. (2016) Why Welch's test is type I error robust. *The Quantitative Methods for Psychology, 12*(1), 30–38. https://www.tqmp.org/RegularArticles/vol12-1/p030/p030.pdf

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Desimone, L. (2011). A primer on effective professional development. *Phi Delta Kappan, 92*(6), 68–71. http://www.jstor.org/stable/25822820

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81–112. http://dx.doi.org/10.3102/01623737024002081

Domin, D. S. (1999). A review of laboratory instruction styles. *Journal of Chemical Education, 76*(4), 543. https://doi.org/10.1021/ed076p543

Doyle, W. (1990). Classroom knowledge as a foundation for teaching. *Teachers College Record, 91*(3), 347–360. https://doi.org/10.1177/016146819009100310

Drits, D. (2011). *Elementary teachers' science practice, beliefs, and content knowledge during and following a reform-based professional development program* (Order No. 3445434) [Doctoral dissertation, University of Utah]. ProQuest Dissertations & Theses Global.

Enochs, L. G., *&* Riggs, I. M. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. *School Science and Mathematics, 90,* 694–706. https://doi.org/10.1111/j.1949-8594.1990.tb12048.x

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403–434. https://doi.org/10.2307/749875

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 465–498. http://heinonline.org.ezproxy.lib.ou.edu/HOL/Page?public=false&handle=hein.journals/hjl28&page=465&collection=journals#

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, *7*(9), 1–23. https://doi.org/10.14507/epaa.v7n9.1999

Fields, E. T., Levy, A. J., Tzur, M. K. Martinez-Gudapakkam, A., & Jablonski, E. (2012). The science of professional development. *Phi Delta Kappan, 93*(8), 44–46. http://dx.doi.org/10.2307/23210372

Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbooks in Economics* (3rd ed.) (pp. 383–421). North–Holland. https://cepa.stanford.edu/sites/default/files/Accountability_Handbook.pdf

Funk, C., & Goo, S. K. (2015). *A look at what the public knows and does not know about science*. Pew Research Center. https://www.pewresearch.org/science/2015/09/10/what-the-public-knows-and-does-not-know-about-science/

Funk, C., & Rainie, L. (2015). *Public and scientists' views on science and society*. Pew Research Center. https://www.pewresearch.org/science/2015/01/29/public-and-scientists-views-on-science-and-society/

Garet, M. S., Birman, B. F., Porter, A. C., Desimone, L., Herman, R., & Yoon, K. S. (1999). *Designing effective professional development: Lessons from the Eisenhower Program* (Report No. ED/OUS99-3). U.S. State Department of Education Office of the Undersecretary. https://files.eric.ed.gov/fulltext/ED442634.pdf
,
Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945. http://www.jstor.org/stable/3202507

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology, 77*(6), 631–645. http://dx.doi.org/10.1037/0022-0663.77.6.631

Grant, S. G., Peterson, P. L., & Shojgreen-Downer, A. (1996). Learning to teach mathematics in the context of system reform. *American Educational Research Journal, 33*(2), 502–541. http://www.jstor.org/stable/1163294

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher, 15*(5), 5–12. http://www.jstor.org/stable/117480

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R. (2009). Closing the knowledge gap on effective professional development. *Educational Horizons, 87*(4), 224–233. http://www.jstor.org/stable/42923773

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500. http://www.jstor.org/stable/20446159

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84–117. http://links.jstor.org/sici?sici=0022-3808%28199202%29100%3A1%3C84%3ATTBCQA%3E2.0.CO%3B2-N

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management, 24*(2), 297–327. https://doi.org/10.1002/pam.20091

Hargreaves, A. (2014). *Handbook of professional development in education: Successful models and practices, PreK–12*. Guilford Publications.

Harlow, D. B. (2013). An investigation of how a physics professional development course influenced the teaching practices of five elementary school teachers. *Journal of Science Teacher Education, 25*, 119–139. http://dx.doi.org/10.1007/s10972-013-9346-z

Harris Interactive. (2011). *Stem perceptions: Study & parent study: Parents and students weigh in on how to inspire the next generation of doctors, scientists, software developers and engineers*. Microsoft Corp. https://news.microsoft.com/download/archived/presskits/citizenship/docs/STEMPerceptionsReport.pdf

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Hauck, N. (2012). *Effects of sustained teacher professional development on the classroom science instruction of elementary school teachers* (Paper no. 1343) [Doctoral dissertation, Utah State University]. Utah State University Digital Commons @ USU, All Graduate Theses and Dissertations.

Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education, 36*(3), 13–15. http://journals.sagepub.com.ezproxy.lib.ou.edu/doi/pdf/10.1177/002248718503600303

Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching, 49*(3), 333–362. https://psycnet.apa.org/doi/10.1002/tea.21004

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25*(4), 12–21. https://doi.org/10.2307/1176776

Holzberger, D., Philipp, A., & Kunter, M. (2020). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 112*(8), 1608–1622. https://doi.org/10.1037/edu0000445

Hourd, S. M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Southwest Educational Development Lab. https://files.eric.ed.gov/fulltext/ED410659.pdf

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking: From childhood to adolescence*. Basic Books.

Institute of Education Sciences (December 2008). *What Works Clearinghouse evidence standards for reviewing studies* (Vers. 1.0). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_version1_standards.pdf

Johnson, C. C., & Fargo, J. D. (2014). A study of the impact of transformative professional development on Hispanic student performance on state mandated assessments of science in elementary school. *Journal of Science Teacher Education, 25*, 845–859. https://doi.org/10.1007/s10972-014-9396-x

Karweit, N. L. (1983). *Time on task: A research review* (Report No. 332). Johns Hopkins University Center for Social Organization of Schools. http://www.eric.ed.gov/PDFS/ED228236.pdf

Kennedy, M. M. (1998). *Form and substance in inservice teacher education* (Research Monograph No. 13). National Institute for Science Education, University of Wisconsin–Madison. Retrieved from https://files.eric.ed.gov/fulltext/ED472719.pdf
.
Knapp, M. S. (1997). Between systematic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research, 67*(2), 227—266. https://doi.org/10.2307/1170627

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Prentice Hall. https://www.fullerton.edu/cice/_resources/pdfs/sl_documents/Experiential%20Learning%20-%20Experience%20A s%20The%20Source%20Of%20Learning%20and%20Development.pdf

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lakshmanan, A., Heath, B. P., Perlmutter, A., & Elder, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching, 48*(5), 534–551. https://doi.org/10.1002/tea.20404

Lederman, N. G., & Lederman, J. S. (2012). Nature of scientific knowledge and scientific inquiry: Building instructional capacity through professional development. In B. J. Fraser et al. (Eds.), *Springer International Handbooks of Education: Vol. 24. Second International Handbook of Science Education* (pp. 335–359). https://link.springer.com/book/10.1007/978-1-4020-9041-7

Lee, C. A., and Houseal, A. (2003). Self-efficacy, standards, and benchmarks as factors in teaching elementary school science. *Journal of Elementary Science Education, 15*(1), 37–55. http://www.jstor.org/stable/43155736
\
Leithwood, K. A., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. The Wallace Foundation. https://wallacefoundation.org/sites/default/files/2023-07/How-Leadership-Influences-Student-Learning.pdf

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. *Phi Delta Kappan, 76*(8), 591–596. http://www.jstor.org/stable/20405409

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15*(2), 129-151. https://doi.org/10.3102/01623737015002129

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics* (2nd ed.). Corwin Press.

Magnusson, S., Borko, H., Krajcik, J. S., & Layman, J. W. (1992). *The relationship between teacher content and pedagogical content knowledge and student content knowledge of heat energy and temperature* [Paper presentation]. National American Association for Research in Science Teaching, Boston, MA.

Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research in Science Teaching, 28*(1), 41–53. https://doi.org/10.1002/tea.3660280105

Marzano, R. J. (2003). *What works in schools: Translating research into action.* Association for Supervision and Curriculum Development.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction: What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474–496. https://doi.org/10.1002/tea.20347

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125–145. https://doi.org/10.1016/0272-7757(94)90003-5

Naghshpour, S. (2016). *A primer on nonparametric analysis* (Volume I). Business Expert Press. https://ebookcentral-proquest.com.ezproxy.lib.ou.edu/lib/ounorman/detail.action?docID=4648712

National Center on Time and Learning. (2010). *The relationship between time and learning: A review of the theoretical research*. http://www.timeandlearning.org/

National Research Council (Ed.). (1996). *National science education standards.* National Academy Press.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (J. D. Bransford, A. L. Brown, & R. R. Cocking, Eds). The National Academies Press.

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

National Research Council. (2013). *Monitoring progress toward successful K–12 STEM education: A nation advancing?* Committee on the Evaluation Framework for Successful K–12 STEM Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. The National Academies Press. https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts

National Science Foundation; Directorate for Education and Human Resources; Division of Elementary, Secondary, and Informal Education. (2010). *MSP Math and Science Partnership Program: Strengthening America by advancing achievement in mathematics and science* (NSF Publication No. 2010-046). National Science Foundation. https://www.nsf.gov/pubs/2005/nsf05069/nsf05069.pdf

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

Office of Educational Quality and Accountability (2015). *Oklahoma school profiles*. http://schoolreportcards.org/report-card/district/14I029/year/2015

Oklahoma Technical Advisory Committee. (2014, August 8). *TAC statement on quality of Oklahoma assessment system (OCCT/EOI/OAAP/WIDA)*. Oklahoma State Department of Education. https://oklahoma.gov/content/dam/ok/en/osde/documents/services/assessments/elementary-and-secondary-education/Attachment-7.pdf

Osborne, J. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Penuel, W. R., Fishman, B. J., Yamaguchi, R., and Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921–958. https://doi.org/10.3102/0002831207308221

Piaget, J. (1977). The development of thought, equilibration of cognitive structures. Oxford: Basil Blackwell.

Posnanski, T. J. (2002). Professional development programs for elementary science teachers: An analysis of teacher self-efficacy beliefs and a professional development model. *Journal of Science Teacher Education, 13*(3), 189–220. http://www.jstor.org/stable/43156296

Powell-Moman, A. D. & Brown-Schild, V B. (2011). The influence of a two-year professional development institute on teacher self-efficacy and use of inquiry-based instruction. *Science Educator*, *20*(2), 47–53. https://files.eric.ed.gov/fulltext/EJ960637.pdf

The President and Fellows of Harvard College, Project MOSART. (2007). *Earth Science Test* (NSF 0412382) [Test form 822]. https://waps.cfa.harvard.edu/mosart/mosart_main/PDFs/5_8_earth.pdf

The President and Fellows of Harvard College, Project MOSART. (2007). *Earth Science Test* (NSF 0412382) [Test form 831]. https://waps.cfa.harvard.edu/mosart/mosart_main/PDFs/9_12_earth.pdf

The President and Fellows of Harvard College, Project MOSART. (2007). *Physical Science Test* (NSF 0412382)) [Test form 721]. https://waps.cfa.harvard.edu/mosart/mosart_main/PDFs/5_8_physical.pdf

The President and Fellows of Harvard College, Project MOSART. (2011). *9–12 Physics* (NSF grant EHR-0830922). https://waps.cfa.harvard.edu/mosart/mosart_main/PDFs/9_12_physics.pdf

The President and Fellows of Harvard College, Project MOSART. (2011). *Life Science Survey Test* (NSF grant EHR-0830922) [Test form 921]. https://waps.cfa.harvard.edu/mosart/mosart_main/PDFs/5_8_life.pdf

The President and Fellows of Harvard College, Project MOSART. (2011). *Life Science Mastery Test for Educators K–12* (NSF grant EHR-0830922). https://waps.cfa.harvard.edu/mosart/mosart_main/PDFs/ls_mastery.pdf

Reading Sufficiency Act, 70 OK Stat § 70-1210.508C (1997)

Rhine, S. (1998). The role of research and teachers’ knowledge base in professional development. *Educational Researcher, 27*(5), 27–31. https://doi.org/10.3102/0013189X027005027

Rhodes, A., & Marder, M. (2024). Measuring the value of teachers from traditional certification pathways in Texas: A comprehensive study. *Education Policy Analysis Archives, 32*(51). https://doi.org/10.14507/epaa.32.8556

Riggs, I. (1988). *The development of an elementary teacher's science efficacy belief instrument* [Doctoral dissertation, Kansas State University]. Kansas State University Proquest Dissertations & Theses.

Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004). *Evaluation: A systematic approach*. SAGE Publications.

Rowland, T., Huckstep, P. & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education, 8*, 255–281. https://doi.org/10.1007/s10857-005-0853-5

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The astronomy and space science concept inventory: Development and validation of assessment instruments aligned with the K–12 national science standards. Astronomy Education Review, 8(1). http://dx.doi.org/10.3847/AER2009024

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. *American Educational Research Journal, 50*(5), 1020-1049. https://doi.org/10.3102/0002831213477

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* University of Tennessee Value-Added Research and Assessment Center. https://www.beteronderwijsnederland.nl/files/cumulative%20and%20residual%20effects%20of%20teachers.pdf

Sandholtz, J. H., & Ringstaff, C. (2014). Inspiring instructional change in elementary school science: The relationship between enhanced self-efficacy and teacher practices. *Journal of Science Teacher Education, 25*(6), 729–751. https://doi.org/10.1007/s10972-014-9393-0

Sawilowsky, S. S. (2009) New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2). https://doi.org/10.22237/jmasm/1257035100

Scott, T. P., Schroeder, C., Tolson, H., Huang, T., & Williams, O. M. (2014). A longitudinal study of a 5th grade science curriculum based on the 5E model. *Science Educator, 23*(1), 49–55. http://search.proquest.com.ezproxy.lib.ou.edu/docview/1552464559/fulltextPDF/8D34CC2EEB184092PQ/1?accountid=12964

Shadish, W. R., Cook, T. D., & Campbell, D. T., (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22. https://doi: 10.17763/haer.57.1.j463w79r56455411

Shulman, L. (1992, September–October). Ways of seeing, ways of knowing, ways of teaching, ways of learning about teaching. *Journal of Curriculum Studies, 28,* 393–396.

Smith, D. C., & Neale, D. C. (1989). The construction of subject matter knowledge in primary science teaching. *Teaching and Teacher Education, 5*(1), 1–20. https://doi.org/10.1016/0742-051X(89)90015-2

Stillman, J., & Schultz, K. (2021). *NEPC Review: 2020 Teacher prep review: Clinical practice and classroom management*. National Education Policy Center. http://nepc.colorado.edu/thinktank/teacher-prep

Stone, E. (2010). *T*-test, independent samples. In *Encyclopedia of research design* (Vol. 0, pp. 1552–1556). SAGE Publications, Inc. https://doi.org/10.4135/ 9781412961288

Stone, E. (2010). *T*-test, paired samples. In *Encyclopedia of research design* (Vol. 0, pp. 1561–1565). SAGE Publications, Inc. https://doi.org/10.4135/9781412961288

Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, *30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy, 14*(3), 331–356 http://dx.doi.org/10.1177/0895904800014003001

Swanson, P. (2010). Teacher efficacy and attrition: Helping students at the introductory levels of language instruction appears critical. *Hispania*, *93*(2), 305-321. https://doi.org/10.1353/hpn.2010.a382877

Tabachnick, B. G., and Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tinnin, R. K. (2000). *The effectiveness of a long-term professional development program on teachers' self-efficacy, attitudes, skills, and knowledge using a thematic learning approach* (UMI No. AAI9992926) [Doctoral dissertation, University of Texas at Austin]. Dissertation Abstracts International: Section A. Humanities and Social Sciences

Traphagen, K. (2011). *Strengthening science education: The power of more time to deepen inquiry and engagement*. National Center on Time and Learning. http://www.timeandlearning.org/files/ StrenghtheningScienceEducation.pdf

Trygstad, P. J., Smith, P. S., Banilower, E. R., & Nelson, M. M. (2013). *The status of elementary science education: Are we ready for the next generation science standards?* Horizon Research, Inc. https://files.eric.ed.gov/fulltext/ED548249.pdf

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

United States Department of Education (1997). *Dwight D. Eisenhower Professional Development Program part B—State and local activities* [CFDA no. 84.164]. https://www2ed.gov/pubs/Biennial/95-96/eval/126-97.pdf

United States Department of Education (2012). *Mathematics and Science Partnerships program highlights: Analytic and technical support for Mathematics and Science Partnerships* [Report no. ED-04-CO-0015]. United States Department of Education. https://www.ed-msp.net/public_ documents/document/policy_ brief/MSP%20Policy%20Brief%20Final.pdf

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE, 8*(7), Article e68967. https://doi.org/10.1371/journal.pone.0068967

VanLone, J., Pansé-Barone, C., & Long, K. (2022). Teacher preparation and the COVID-19 disruption: Understanding the impact and implications for novice teachers. *International Journeal of Educational Research, 3*. https://doi.org/10.1016/j.ijedro.2021.100120

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91. https://doi.org/10.1016/j.tate.2007.01.004

Vygotsky, L. (1934). *Thought and language*. MIT Press.

Vygotsky, L. (1978) *Mind in society: The development of higher psychological processes.* Harvard University Press.

Ward-Horner, J., & Sturmey, P. (2013). Component analyses using single-subject experimental designs: A review. *Journal of Applied Behavior Analysis, 43*(4), 685–704. https://doi.org/10.1901/jaba.2010.43-685

,

Weiss, I. R., Montgomery, D. L., Ridgway, C. J., & Bond, S. L. (1998). *Highlights of the local systemic change through teacher enhancement year three cross-site report.* Horizon Research, Inc. https://www.ntnu.no/wiki/download/attachments/11273030/Weiss,+I.R.+(1998)_Highlights+of+the+LSC+through+Teacher+Enhancement.pdf

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67. https://psycnet.apa.org/doi/10.1023/A:1007999204543

Yoon, K. S., Duncan, T., Lee, S. W-Y., Scarloss, B., and Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, No. 033). Regional Educational Laboratory Southwest. https://files.eric.ed.gov/fulltext/ED498548.

Zeichner, K. M., Darling-Hammond, L., Berman, A. I., Dong, D., & Sykes, G. (2024). *Evaluating and improving teacher teacher preparation programs: Consensus report*. National Academy of Education. https://naeducation.org/wp-content/uploads/2025/01/Evaluating_and_Improving_Teacher_Preparation_Programs.pdf?utm_source=chatgpt.com

Zweip, S. G., & Benken, B. M. (2012). Exploring teachers' knowledge and perceptions across mathematics and science through content-rich learning experiences in a professional development setting. *International Journal of Science and Mathematics Education, 11*, 299–324. http://dx.doi.org/10.1007/s10763-012-9334-3

Appendix A: Middle School Teacher Interview Questions and Responses

Responses to: As a middle school teacher, are you satisfied with the science background that students have when they enter your class? Why or why not?

- No, their background is unsatisfactory. At the 7th grade level we spend much of our time covering material that should've been mastered at the elementary level rather than extending the [science] skills and standards at an appropriate higher level.
- I am not satisfied as a science teacher but as [an] educator with many years of experience I do understand the problems facing the elementary teachers. Training that helps foster an understanding as to how the curriculum can be integrated, meeting all standards, would be very beneficial.
- Not satisfied because they have a very limited background.
- No, I feel that students could be far better prepared for middle school science. Students have such limited understanding I am the beginning of their science education.
- NO. Student do not come even with the basics of any science. We have to start from ground level in middle school.
- No, during middle school, our focus should be to continue to develop their process skills and teach them content that will prepare them to be successful at the next level. Instead, as middle school instructors we are often forced to introduce skills, concepts, and content that the students should have already come to our classes with from their elementary science programs. Therefore, the building up and development of skills does not happen and they leave middle school unprepared for the rigors of high school biology, chemistry, and physics. I do not blame the elementary teachers; their instructional time is unequally balanced toward English and math skills. I wholeheartedly believe in strong English and math skills in order

to be successful in science, but I wonder if science could be integrated into the math and English units at the elementary level. Instead of having specific "science time," science could be taught continuously as an integrated subject while teaching English and math.

- Most haven’t even had a science class and we start with a blank slate.
- No, they seem to struggle over basic experimental design.
- A small group of my sixth graders had a really excellent science education from a teacher somewhere in their elementary school experience. They know the parts of an experiment, have a keen sense of the purpose and flow of experiments, they understand underlying and overarching principles of science. This is NOT the norm, because the overwhelming majority have no experiment sense, or idea of underlying concepts that connect sciences together.
- Students are not prepared for science and are not ready in sixth grade. School districts and the state are not building in time for science and focus on math and reading remediation.
- Students are not ready for middle school. Students are mostly prepared for inquiry by the eighth grade but lack test-taking skills to help them analyze data, etc.
- I teach seventh grade, and I feel most of them have a pretty good background by the time they get to me.
- Not really. It depends on what elementary school they went to. I think because of the push to excel in reading and math, something has to give, and science is usually that something.
- I know that elementary teachers are pressured by the state to make certain scores so I cannot entirely blame primary schools. Some kids go to schools that do a lot more science than others so that can be frustrating to everyone.
- NO, They don't know how to set up an experiment, evaluate if their answer makes sense and communicate their results through writing or graphical representation.

- No, not at all. They should be coming in able to use equipment and to understand basic process skills. This is usually because science is not being taught at the elementary level

Responses to: Do you believe the elementary science curriculum is closely aligned with the middle school science curriculum?

- Several students and parents complain that 6th grade is their first experience with science lessons.
- Most students need to experience more hands on learning.
- Not sure if any science is done in elementary. The focus is on reading and math.
- We are told that at the elementary feeder school to our middle school that they do not have time to teach science at all. That their students can't read and the extra time that would be used to teach science is used to teach reading.
- Students do not seem to know the basic process skills needed to be successful in middle school science.
- As a 6th grade teacher, my students come to me with a wide range of prior science knowledge. About 1/5 of my students in each class have understanding of basic science principles, understanding of the scientific method or the nature of an experiment. The rest of the students have little or no experience or knowledge of science. It would be great to have students come to middle school with more of a uniform experience. Somewhere in our feeder schools, there is one school or a teacher teaching some great science, but not at any of the other schools. My guess is that the emphasis is solely on reading and math, but truly...reading and math can be emphasized through quality science teaching.
- Most students do not have any science unless it is an elective or enrichment for gifted in talented. There is not time built into the schedule for science. The focus is on reading and

math. Many elementary teachers do not feel comfortable teaching science and look at it as one more thing on their plate to have to do. They do not have basic scientific equipment available to teach skills nor do they understand how the inquiry system works. It is very concerning that most students haven't performed a true scientific inquiry until 6th grade which makes it very difficult for students and teachers to meet the demands of testing. There is a huge need for funding, support and training at the elementary level. The state department needs to help teachers, not just the district.

- The curriculum may be aligned, but I'm not sure about actual teaching practices.
- I don't think the elementary schools spend much time on science at all.
- I will honestly say that I do need to take another look at the elementary curriculum. I feel that elementary is more geared towards reading, writing, spelling, grammar, and math. While all of those are extremely important students need to be challenged with inquiry. They need to be able to solve a problem with the items placed in front of them. They need to be able to use their hands and create a project.
- They rarely have science class in the elementary schools.
- We are not seeing this for all students coming into middle school.

Responses to: Do students entering middle school science classes understand how the use of math supports the work of science?

- The connection has not been made yet that math and science are closely related.
- Math skills are extremely poor when they get to middle school.
- Students do not see any connection between math and science.
- Students have never had any science.
- Students have trouble understanding the correlation between math and science.

- They seem to be overwhelmed by the presentation of formulas for density, measuring volume or calculating mass on a triple beam balance.
- Most students really don’t see this connection at all.
- I am consistently reminding students that math and science work together. I am always asked at the beginning of the year why do we have to do math in science class. Once we get into a routine they do better understanding that we are always going to use math in science.
- Students do not understand that math skills are essential to understanding much of the science.

Appendix B: Percent Passing—Fifth-Grade Science OCCT

Table B1

Percent Scoring Proficient or Above on Fifth-Grade Science OCCT by District and Demographics

	District 1	District 2	District 3	District 4	District 5
Race and Ethnicity					
Overall	60	68	64	57	49
Hispanic/Latino	59	60	62	39	55
American Indian/Alaskan Native	55	89	65	42	71
Asian	n/a	92	82	87	79
Black/African American	53	45	61	39	38
Pacific Islander	n/a	75	100	0	33
White/Caucasian	67	78	74	69	65
Two or more races	72	74	72	71	53
Gender					
Female	67	72	71	63	53
Male	65	75	72	66	60
Socioeconomics					
Economically disadvantaged	55	54	59	50	49
Non-economically disadvantaged	72	80	81	75	71
English Language Proficiency					
English Language Learners (ELL)	0	29	36	23	19
ELL first year proficient	n/a	14	70	42	46
ELL second year proficient	n/a	63	49	54	58
Non-English language learners (non-ELL)	60	75	66	59	52
Students with Individualized Education Programs (IEPs)					
IEP with accommodations	26	29	25	24	23
IEP without accommodations	70	51	53	44	47
Individualized Education Program (IEP)	33	38	34	30	29
Regular education	66	74	72	64	56

Table B2

Fifth-Grade Science OCCT Scores Median Percent Correct by Process Standards and Objectives

Standards	P1.0	P1.1	P1.2	P2.0	P2.1	P2.2	P3.0	P3.2	P3.4	P4.0	P4.2	P4.3	P4.4
District 1													
Regular Education	84	85	84	70	64	74	80	75	91	78	85	72	74
ELL	67	50	80	44	25	60	77	67	100	57	83	25	50
Non-ELL	82	82	82	68	63	72	78	73	90	76	83	70	70
IEP	64	55	71	56	54	60	64	58	85	60	73	56	51
District 2													
Regular Education	87	88	86	73	68	77	82	77	92	82	86	77	80
ELL	69	65	69	52	53	58	55	52	70	58	68	53	50
Non-ELL	86	87	85	72	67	75	81	76	91	81	85	76	78
IEP	73	72	73	59	55	60	68	62	82	65	76	59	61
District 3													
Regular Education	85	86	84	74	67	80	83	77	94	82	85	80	77
ELL	70	60	72	57	53	60	59	58	77	68	75	65	54
Non-ELL	83	84	83	72	65	77	81	74	93	80	83	77	74
IEP	69	68	71	58	52	63	66	56	87	64	73	61	56
District 4													
Regular Education	82	85	80	69	65	71	80	74	92	78	85	70	74
ELL	65	66	65	57	56	56	57	48	81	63	73	55	49
Non-ELL	80	83	78	67	63	69	78	72	91	75	83	67	74
IEP	68	67	69	57	55	58	63	54	83	61	73	53	52
District 5													
Regular Education	80	80	81	67	62	71	76	70	90	74	81	67	70
ELL	61	60	64	53	50	54	59	52	77	57	71	50	49
Non-ELL	79	78	79	65	60	69	74	68	89	72	80	64	67
IEP	66	63	68	55	52	58	62	53	84	60	72	54	53

Table B3

Fifth-Grade Science OCCT Scores Median Percent Correct by Content Standards and Objectives

Standards	C1.0	C1.1	C1.2	C1.3	C1.4	C2.0	C2.1	C2.2	C3.0	C3.1	C3.2	C3.3
District 1												
Regular Education	76	78	86	75	63	78	82	75	77	81	72	79
ELL	67	60	60	100	50	60	40	80	46	75	20	50
Non-ELL	74	77	83	71	61	77	80	73	74	79	69	77
IEP	60	69	66	52	45	77	80	73	74	79	69	77
District 2												
Regular Education	79	82	88	77	66	82	83	81	80	82	76	83
ELL	56	62	70	44	38	61	60	62	59	65	55	61
Non-ELL	78	81	87	75	65	81	82	79	78	81	74	82
IEP	64	68	77	58	52	68	70	65	64	68	61	65
District 3												
Regular Education	80	81	87	76	74	82	82	81	77	79	74	80
ELL	61	71	65	56	63	66	61	71	62	69	60	53
Non-ELL	78	79	85	73	70	80	80	79	75	78	72	77
IEP	62	69	71	57	48	65	66	64	61	38	58	58
District 4												
Regular Education	76	79	85	75	59	77	79	74	76	76	72	80
ELL	57	65	67	49	42	63	62	65	55	55	53	58
Non-ELL	74	77	83	71	57	75	76	72	73	73	70	78
IEP	60	67	67	52	45	62	64	61	58	61	56	60
District 5												
Regular Education	72	74	83	70	60	75	75	73	72	76	68	75
ELL	57	60	69	59	41	58	55	59	62	55	49	52
Non-ELL	71	73	81	68	58	73	73	72	70	73	66	72
IEP	60	67	68	56	44	62	60	62	57	63	55	58

Note: Includes students enrolled for the Full Academic Year (FAY) only. OCCT = Oklahoma Core Curriculum Test.

Appendix C: MOSART[1] Needs Assessment Average Scores per Topic

Domain	Topics (from *National Science Education Standards* and *Benchmarks for Science Literacy)*	Percent Correct (*n* = 15)
Space Science	Gravity is the force that keeps planets in orbit around the sun and governs the rest of the motion in the solar system. Gravity alone holds us to the Earth's surface and explains the phenomenon of the tides.	20%
Space Science	The sun is the major source of energy for phenomena on the Earth's surface, such as growth of plants, winds, ocean currents, and the water cycle. Seasons result from variations in the amount of the sun's energy hitting the surface, due to the tilt of the Earth's rotation on its axis and the length of the day.	66.7%
Space Science	The Earth is the third planet from the sun in a system that includes the moon, the sun, eight other planets and their moons, and smaller objects, such as asteroids and comets. The sun, an average star, is the central and largest body in the solar system.	60%
Space Science	The sun is many thousands of times closer to the Earth than any other star. Light from the sun takes a few minutes to reach the Earth, but light form the next nearest star takes a few years to arrive. The trip to that star would take the fastest rocket thousands of years. Some distant galaxies are so far away that their light takes several billion years to reach the Earth. People on Earth, therefore, see them as they were that long ago in the past.	26.7%
	Space Science Overall	**43.3%**
Life Science	All organisms are composed of cells-the fundamental unit of life. Most organisms are single cells; other organisms, including humans, are multicellular.	86.7%
Life Science	Specialized cells perform specialized functions in multicellular organisms. Groups of specialized cells cooperate to form a tissue, such as a muscle. Different tissues re in turn grouped together to form larger functional units, called organs. Each type of cell, tissue, and organ has a distinct structure and set of functions that serve the organism as a whole.	56.7%
Life Science	Regulation of an organism's internal environment involves sensing the internal environment and changing physiological activities to keep conditions within the range required to survive.	53.3%
Life Science	All organisms must be able to obtain and use resources, grow, reproduce, and maintain stable internal conditions while living in a constantly changing external environment.	80%
Life Science	Disease is a breakdown in structures or functions of an organism. Some diseases are the result of intrinsic failures of the system. Others are the result of damage by infection by other organisms.	80%

Life Science	Biological evolution accounts for the diversity of species developed through gradual processes over many generations. Species acquire many of their unique characteristics through biological adaptation, which involves the selection of naturally occurring variations in populations. Biological adaptations include changes in structures, behaviors, or physiology that enhance survival and reproductive success in a particular environment.	73.3%
Life Science	Cells carry on the many functions needed to sustain life. They grow and divide, thereby producing more cells. This requires that they take in nutrients, which they use to provide energy for the work that cells do and to make the materials that a cell or an organism needs.	60%
Life Science	Behavior is one kind of response an organism can make to an internal or environmental stimulus. A behavioral response requires coordination and communication at many levels, including cells, organ systems, and whole organisms. Behavioral response is a set of actions determined in part by heredity and in part from experience.	86.7%
Life Science	The number of organisms an ecosystem can support depends on the resources available and abiotic factors, such as quantity of light and water, range of temperatures, and soil composition. Given adequate biotic and abiotic resources and no disease or predators, populations (including humans) increase at rapid rates. Lack of resources and other factors, such as predation and climate, limit the growth of the populations in specific niches in the ecosystem.	53.3%
Life Science	Every organism requires a set of instructions for specifying traits. Heredity is the passage of these instructions from one generation to another.	0%
Life Science	An organism's behavior evolves through adaptation to its environment. How a species moves, obtains food, reproduces, and responds to danger are based in the species' evolutionary history.	73.3%
Life Science	The characteristics of an organism can be described in terms of a combination of traits. Some traits are inherited from others and others result from interactions with the environment.	46.7%
Life Science	Living systems at all levels of organization demonstrate the complementary nature of structure and function. Important levels of organization for structure and function include cells, organs, tissues, organ systems, whole organisms, and ecosystems.	73.3%
	Life Science Overall	**63.3%**
Physical Science	Chemical elements do not break down during normal laboratory reactions involving such treatment as heating, exposure to electric current, or reaction with acids. There are more than 100 known elements that combine in a multitude of ways to produce compounds, which account for the living and nonliving substances that we encounter.	86.7%
Physical Science	Heat moves in predictable ways, flowing from warmer objects to cooler ones, until both reach the same temperature.	20%
Physical Science	Electrical circuits provide a means of transferring electrical energy when heat, light, sound, and chemical changes are produced.	33.3%

Physical Science	The sun is a major source of energy for changes on the Earth's surface. The sun loses energy by emitting light. A tiny fraction of that light reaches the Earth, transferring energy from the sun to the Earth. The sun's energy arrives as light with a range of wavelengths, consisting of visible light, infrared, and ultraviolet radiation.	46.7%
Physical Science	Energy is a property of many substances and is associated with heat, light, electricity, mechanical motion, sound, nuclei, and the nature of a chemical. Energy is transferred in many ways.	60%
Physical Science	In most chemical and nuclear reactions, energy is transferred into or out of a system. Heat, light, mechanical motion, or electricity might all be involved in such transfers.	13.3%
Physical Science	Light interacts with matter by transmission (including refraction), absorption, or scattering (including reflection). To see an object, light from that object—emitted by or scattered from it—must enter the eye.	46.7%
Physical Science	A substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the sample. A mixture of substances often can be separated into the original substances using one or more of the characteristic properties.	53.3%
	Physical Science Overall	**45%**
Earth Science	Fossils provide important evidence of how life and environmental conditions have changed.	93.3%
Earth Science	Landforms are the result of a combination of constructive and destructive forces. Constructive forces include crustal deformation, volcanic eruption, and deposition of sediment, while destructive forces include weathering and erosion.	66.7%
Earth Science	Water, which covers the majority of the Earth's surface, circulates through the crust, oceans, and atmosphere in what is known as the "water cycle". Water evaporates from the Earth's surface, rises and cools as it moves to higher elevations, condenses as rain or snow, and falls to the surface where it collects in lakes, oceans, soil, and in rocks underground.	66.7%
Earth Science	Water is a solvent. As it passes through the water cycle it dissolves minerals and gases and carries them to the oceans.	73.3%
Earth Science	The Earth processes we see today, including erosion, movement of lithospheric plates, and changes in atmospheric composition, are similar to those that occurred in the past. Earth history is also influenced by occasional catastrophes, such as the impact of an asteroid or comet.	20%
Earth Science	The solid Earth is layered with a lithosphere; hot, convecting mantle, and dense, metallic core.	80%
Earth Science	The atmosphere is a mixture of nitrogen, oxygen, and trace gases that include water vapor. The atmosphere has different properties at different elevations.	93.3%
Earth Science	Some changes in the solid Earth can be described as the "rock cycle". Old rocks at the Earth's surface weather, forming sediments that are buried, then compacted, heated, and often recrystallized into new rock.	80%

	Eventually, those new rocks may be brought to the surface by the forces that drive plate motions, and the rock cycle continues.	
Earth Science	Lithospheric plates on the scales of continents and oceans constantly move at rates of centimeters per year in response to movements in the mantle. Major geological events, such as earthquakes, volcanic eruptions, and mountain building, result from these plate motions.	80%
	Earth Science Overall	**72.6%**
	OVERALL AVERAGE	**62.9%**

Appendix D: Summer Institute Activities

During the summer institute (SI), STEM faculty engaged teachers in inquiry-based integrated science-math lessons aimed at increasing teachers' content knowledge while simultaneously modeling sound teaching strategies. SI lessons were purposefully planned in a modified 5E format to include the integration of engineering, math, and science disciplinary core ideas and practices. Each lesson consisted of the following parts:

- **Engage**- Participants completed a formative assessment probe to elicit what they already understood about the topic. This was designed to help participants understand the gaps in their understanding.
- **Explore**- Participants engaged in an activity related to the topic that was geared toward secondary and/or college students. This was designed to help participants understand the concepts at a level beyond that which they teach.
- **Explain**- Expert scientists led a discussion/explanation of the concepts introduced during the explore activity.
- **Extend**- Participants engaged in an activity related to the topic that was on-level for the grade which the participants teach. This was designed to help to reinforce the concepts learned from the explore and explain phases, but also model for the teachers how the topic could be applied in their own classrooms.
- **Evaluate**- Participants revisited the formative assessment probe introduced in the engage portion of the lesson. This time, they were expected to write a formal scientific argument using the concepts that they learned throughout the lesson as a response to the probe.

The activities that were employed during the Summer Institute are as follows:

	Strand: Physical Science (Energy)	Strand: Life Science (Interdependence and Biodiversity)	Strand: Earth Science (Relative Dating)
Engage	Formative Assessment Probe	Formative Assessment Probe	Formative Assessment Probe
Explore	Build an Electromagnet Build a Windmill	Bug Lyphe	Model of 3 Faults Geologic Time
Explain	Learn from an Expert	Learn from an Expert	Learn from an Expert
Extend	Catching the Wind An Alarming Idea	Best of Bugs	A Stick in the Mud
Evaluate	Formative Assessment Probe	Formative Assessment Probe	Formative Assessment Probe

Detailed descriptions of each lesson and activity are provided in the section that follows.

Pretests

Description: Participants completed formative assessment probes to elicit what participants already understood about the topics. This helped participants to become aware of the gaps in their understanding. Later, participants revisited the formative assessment probes in order to write formal scientific arguments using the concepts that they learned throughout the lesson as a response to the probes. The probes were selected because they aligned with the specific activities in which participants were to engage. In addition, only probes that could elicit scientific argumentation were selected. Scientific argumentation is a skill that is promoted by the Oklahoma Academic Standards for Science (OAS-S); however, few elementary teachers have experience developing scientific arguments. The use of these probes were intended to prepare teachers for implementation of OAS-S, while also eliciting participants' understanding of the concepts. The specific probes used were titled "Mountaintop Fossil," "Mountain Age," "How Can You Make an Electromagnet Stronger," and "How Can You Make an Electromagnet."

Session leader: K20 Center Elementary STEM Specialist

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- *Uncovering Student Ideas in Science: 25 Formative Assessment Probes (Vol. 1)* by Keeley, Eberle, and Farrin
- *Uncovering Student Ideas in Science: 25 More Formative Assessment Probes (Vol. 2)* by Keeley, Eberle, and Tugel
- *Uncovering Student Ideas in Physical Science, Volume 2* by Keeley and Harrington
- Large sticky chart paper
- Markers
- Copy paper
- Pens

Standards:

- 4th grade (OAS): 4-PS3-1, 4-PS3-2, 4-PS3-4, 4-LS1-1, 4-ESS1-1, 4-ESS2-2,
- 5th grade (PASS): C1.3, C3.1
- Beyond 5th grade: MS-PS2-3, MS-LS4-1, MS-ESS1-4, MS-ESS2-2, MS-ESS2-3, HS-PS2-5

Build an Electromagnet

Description: Participants built an electromagnet. They investigated the effects of the number of batteries, the types of circuits, and the number of coils on the strength of the electromagnet. The activity was designed to teach high school physical science content. This activity acted as a precursor for the later “Build a Windmill” activity.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea “Conservation of Energy and Energy Transfer.” Student Oklahoma Core Curriculum Test (OCCT) science scores were low in this area (58%), and elementary teacher MOSART scores were low in this area (69.9%).

Session leader: St. Gregory’s University Science Education Professor, Oklahoma City Community College Physics Professor

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials

- Pencils
- Thin, insulated wire
- D batteries
- Paperclips

Standards: MS-PS2-3, HS-PS2-5

Build a Windmill

Description: Participants designed and constructed a windmill, and tested and revised their constructions. Participants learned how windmills transform wind energy into electrical energy. The activity was designed to teach high school physical science content.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea "Conservation of Energy and Energy Transfer." Student OCCT science scores were low in this area (58%), and elementary teacher MOSART scores were low in this area (69.9%).

Session leader: St. Gregory's University Science Education Professor, Oklahoma City Community College Physics Professor

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- Hairdryers
- Small objects for each team to lift
- Wooden sticks, wooden spoons, small wooden (balsa) pieces, bendable wire, string, paperclips, rubber bands, toothpicks, aluminum foil, tape, dowels, glue, paper, cardboard, plastic wrap

Standards: HS-PS3-3, HS-ESS3-2, HS-ESS3-4

Learn from an Expert: Conservation of Energy and Energy Transfer

Description: A content expert on this topic and a scientist who studies this topic led participants in an interactive discussion intended for participants to gain a better understanding about this topic.

Rationale: Data indicated that elementary teachers struggle with the disciplinary core idea "Conservation of Energy and Energy Transfer." Student OCCT science scores were low in this area (58%), and elementary teacher MOSART scores were low in this area (69.9%).

Session leaders: University of Oklahoma Professor of Geology and Geophysics, St. Gregory's University Science Education Professor, Oklahoma City Community College Physics Professor

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Standards:

- 4th grade (OAS): 4-PS3-1, 4-PS3-2, 4-PS3-4, 4-LS1-1, 4-ESS1-1, 4-ESS2-2,
- 5th grade (PASS): C1.3, C3.1
- Beyond 5th grade: MS-PS2-3, MS-LS4-1, MS-ESS1-4, MS-ESS2-2, MS-ESS2-3, HS-PS2-5, HS-PS3-3, HS-ESS3-2, HS-ESS3-4

Catching the Wind

Description: Participants engaged in an activity related to the topic that is on-level for the grade which the participants teach. This helped to reinforce the concepts learned from the explore and explain phases, but also modeled for the teachers how the topic can be applied in their own classrooms. In this activity, participants thought like mechanical engineers—and also used their understanding of air as wind—to design and create wind-powered machines. The storybook *Leif Catches the Wind* introduced participants to wind turbines that generate renewable energy. Participants explored how common machines such as mechanical pencils and egg beaters work, then used their mechanical engineering skills to design sailboats and windmills that catch the wind.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea “Conservation of Energy and Energy Transfer”. Student OCCT science scores were low in this area (58%), and elementary teacher MOSART scores were low in this area (69.9%).

Session leader: K20 Center Elementary STEM Specialist

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- Engineering is Elementary (EiE) Catching the Wind: Designing Windmills kit

Standards:

- OAS-S: 4-PS3-1, 4-PS3-2, 4-PS3-4, 4-ESS3-1
- 5th grade PASS: C1.3, C2.2a

An Alarming Idea

Description: Participants engaged in an activity related to the topic that is on-level for the grade which the participants teach. This helped to reinforce the concepts learned from the explore and explain phases, but also modeled for the teachers how the topic could be applied in their own classrooms. The lessons in this unit were chosen to get participants thinking like electrical engineers, starting with the storybook *A Reminder for Emily*, about a girl living in the Australian outback who needs an alarm to remind her when it’s time for chores around the ranch. The hands-on activities in this unit reinforced science concepts including conductors and insulators, schematic diagrams, and circuits. For the final engineering design challenge, students planned, created, and improved their own alarm circuit.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea "Conservation of Energy and Energy Transfer." Student OCCT science scores were low in this area (58%), and elementary teacher MOSART scores were low in this area (69.9%).

Session leader: K20 Center Elementary STEM Specialist

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:
- Engineering is Elementary (EiE) An Alarming Idea: Designing Alarm Circuits kit

Standards:
- OAS-S: 4-PS3-4, 5-PS1-3
- 5th grade PASS: C1.3

Learn from an Expert: Argumentation in Science

Description: Participants learned about the Claims–Evidence–Reasoning (CER) model for writing scientific arguments, and practiced writing their own CERs.

Rationale: Scientific argumentation is a skill that is promoted by the Oklahoma Academic Standards for Science (OAS-S); however, few elementary teachers have experience developing scientific arguments. This activity helped prepare teachers for implementation of OAS-S.

Session leader: Partner district science content specialist

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:
- Large sticky paper
- Markers

Standards: OAS Science and Engineering Practice #7: Engaging in argument from evidence

Learn from an Expert: Graphic Organizers

Description: Participants learned how to create effective graphic organizers for science class. Participants then developed graphic organizers to accompany “Catching the Wind” and “An Alarming Idea”.

Rationale: OCCT data suggest that ELL students and students on IEPs struggled in science. Graphic organizers are one differentiation strategy that has been shown to be effective in promoting success among these students in these disciplines.

Session leader: Partner district science specialists with endorsements in ELL pedagogy

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- Writing paper
- Pens

Bug Lyphe

Description: Participants designed and set traps for insects in various locations. They observed and statistically analyzed the insects trapped, and explained why insects with certain adaptations were found in specific areas. This lesson was designed for grades 6–12.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core ideas “Structure and Function” (SF) and “Interdependent Relationships in Ecosystems” (IRE). Student OCCT science scores were low in IRE (62%), and elementary teacher MOSART scores were low in these areas (SF=66.7%, IRE=65%). In addition, SF is a topic that 4th grade teachers had not taught previously under PASS. Since it was a new topic for teachers, they likely had not had recent experiences with the topic.

Session leader: University of Oklahoma Professor of Botany and Microbiology

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- Sweep nets
- Plastic cups
- Sticky traps
- Forceps

Standards: MS-LS2-1, HS-LS2-1, HS-LS2-2, HS-LS2-6, 5th grade C2.1

Learn from an Expert: Interdependence and Biodiversity

Description: A content expert on this topic and a scientist who studies this topic led participants in an interactive discussion so that participants could gain a better understanding about this topic.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core ideas “Structure and Function” (SF) and “Interdependent Relationships in Ecosystems” (IRE). Student OCCT science scores were low in IRE (62%), and elementary teacher MOSART scores were low in these areas (SF=66.7%, IRE=65%). In addition, SF was a topic that 4th grade teachers had not taught previously under PASS. Since it was a new topic for teachers, they likely had not had recent experiences with the topic.

Session leader: University of Oklahoma Professor of Botany and Microbiology

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Best of Bugs

Description: Participants engaged in an activity related to the topic that is on-level for the grade which the participants teach. This helped to reinforce the concepts learned from the explore and explain phases, but also modeled for the teachers how the topic could be applied in their own classrooms. In this activity, participants became agricultural engineers. They applied their knowledge of insects, insect life cycles, pollination, and natural systems as they tested a variety of materials, then engineered their own technologies for pollinating plants by hand.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core ideas “Structure and Function” (SF) and “Interdependent Relationships in Ecosystems” (IRE). Student OCCT science scores were low in IRE (62%), and elementary teacher MOSART scores were low in these areas (SF=66.7%, IRE=65%). In addition, SF was a topic that 4th grade teachers had not taught previously under PASS. Since it was a new topic for teachers, they likely had not had recent experiences with the topic.

Session leader: K20 Center Elementary STEM Specialist

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials: Engineering is Elementary (EiE) The Best of Bugs: Designing Hand Pollinators kit

Standards:

- 4-LS1-1
- 5th grade C2.1

Learn from an Expert: Cooperative Learning

Description: Participants learned how to facilitate effective cooperative learning strategies for science. Participants then developed cooperative learning strategies to accompany “Best of Bugs.”

Rationale: OCCT data suggested that ELL students and students on IEPs struggled in science. Cooperative learning is one differentiation strategy that has been shown to be effective in promoting success among these students in these disciplines.

Session leader: Partner district science specialists with endorsements in ELL pedagogy

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Model of 3 Faults

Description: Participants engaged in an activity in which they created models of fault movements of the Earth's surface. This lesson was intended for grades 7–12.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea “Plate Tectonics and Large-Scale System Interactions.” Student OCCT science scores were low in this area (54%).

Session leader: University of Oklahoma Professor of Geology and Geophysics

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- Physiographic map of the world
- Crayons or colored pencils
- Scissors
- Tape or glue
- Metric ruler
- Construction paper

Standards:

- MS-ESS2-2
- HS-ESS2-1
- HS-ESS2-3

Geologic Time

Description: Participants engaged in activities that led them to understand how scientists use features found in rock layers to relatively date the age of the sample. This lesson was intended for grades 7–12.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea “Plate Tectonics and Large-Scale System Interactions”. Student OCCT science scores were low in this area (54%).

Session leader: University of Oklahoma Professor of Geology and Geophysics

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials:

- Playing cards
- Graduated cylinders
- Several different types of sand
- Funnels
- Clear plastic cups
- 2- 5lb bags of sediment
- Fossils: trilobite, human, brachiopod, fish, reptile, amphibian, flower, dinosaur, mammal
- Rolls of adding machine tape
- Scotch tape
- Meter sticks

Standards:

- MS-LS4-1
- MS-ESS1-4
- MS-ESS2-2
- MS-ESS2-3
- HS-ESS1-5

Learn from an Expert: Relative Dating

Description: A content expert and scientist who studies this topic led participants in an interactive discussion so that participants could gain a better understanding about this topic.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea “Plate Tectonics and Large-Scale System Interactions”. Student OCCT science scores were low in this area (54%).

Session leader: University of Oklahoma Professor of Geology and Geophysics

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

A Stick in the Mud

Description: Participants engaged in an activity related to the topic that is on-level for the grade which the participants teach. This helped to reinforce the concepts learned from the explore and explain phases, but also modeled for the teachers how the topic could be applied in their own classrooms. In this activity, participants dug into the role of geotechnical engineers in order to select a safe, flood-proof, and erosion-proof location for a new TarPul. Working with a model riverbank, they study soil properties, examine maps to assess the potential for erosion at different sites along the river, and factor in the villagers’ preferences for a TarPul location.

Rationale: Data indicated that elementary teachers struggled with the disciplinary core idea “Plate Tectonics and Large-Scale System Interactions”. Student OCCT science scores were low in this area (54%).

Session leader: K20 Center Elementary STEM Specialist

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

Materials: Engineering is Elementary (EiE) A Stick in the Mud: Evaluating a Landscape kit

Standards:

- 4-ESS2-1
- 4-ESS1-1
- 4-ESS3-2
- 5th grade C3.1

Learn from an Expert: Word Walls

Description: Participants learned how to create effective word walls for science. Participants then developed a word wall template to accompany "A Stick in the Mud".

Rationale: OCCT data suggested that ELL students and students on IEPs struggled in math and science. Word walls are one differentiation strategy that has been shown to be effective in promoting success among these students in these disciplines.

Session leader: Partner district science specialists with endorsements in ELL pedagogy

Session facilitators: Partner district science specialists facilitated by managing supplies and providing assistance to the leader and participants as needed.

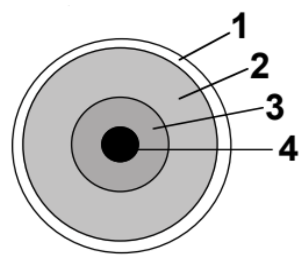
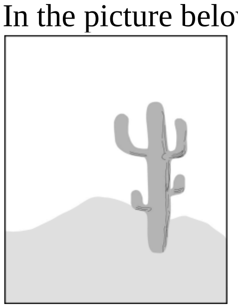
Appendix E: Content Priority List

Content Area	Core Ideas	Grade	Correlation to PASS	Average OCCT Score	STEBI-A Problem Numbers	Average MOSART Score	Comments
ESS	Plate Tectonics and Large-Scale System Interactions	4	5th- C3.2	54%	ESS13	80	Low OCCT
ESS	Earth and the Solar System	5	5th- C3.3	59%	SS1, SS2	43.35	Low OCCT; Low MOSART
ESS	Earth Materials and Systems	4, 5	4th- C4.1; 5th- C3.1; 5th- C3.2	59%	ESS6, ESS7, ESS 8, ESS10, ESS12	73.34	Low OCCT; Content present in both grades
ESS	The Universe and Its Stars	5	4th- C4.4		SS3, SS4	43.34	Low MOSART
ESS	History of Planet Earth	4	4th- C4.3		ESS5, ESS9	43.35	Low MOSART
ESS	Natural Hazards	4					New to OAS
ESS	Roles of Water in Earth's Surface Processes	5	4th- C4.2; 5th- C3.2	54%			Low OCCT
LS	Structure and Function	4			LS1, LS4, LS9, LS15	66.66	Low MOSART; new to OAS
LS	Organization of Matter and Energy Flow in Organisms	5	4th- C3.4; 5th- C2.1	62%			Low OCCT; Content present in both grades
LS	Interdependent Relationships in Ecosystems	5	4th- C3.1; 5th- C2.1	62%	LS3, LS6, LS10, LS12	64.98	Low OCCT; Low MOSART; Content present in both grades
LS	Cycles of Matter and Energy Transfer in Ecosystems	5	4th- C3.4; 5th- C2.1	62%			Low OCCT; Content present in both grades
LS	Information Processing	4					New content
PS	Definitions of Energy	4	4th- C2.1; 5th- C1.4	47%			Low OCCT; Content present of both grades
PS	Energy in Chemical Processes and Everyday Life	4, 5			LS2, LS5, LS8	68.87	Low MOSART; new to OAS
PS	Conservation of Energy and Energy Transfer	4	5th- C1.3	58%	PS2, PS3, PS6, PS7	31.65	Low OCCT; Low MOSART; Content present in both grades

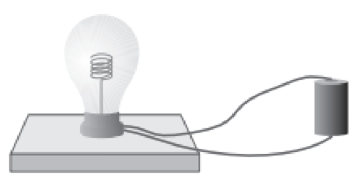
PS	Structure and Properties of Matter	5	5th- C1.1	66%	PS1, PS9	69.99	Low OCCT; Low MOSART; Content present in both grades
PS	Relationship between Energy and Forces	4					New to OAS
PS	Electromagnetic Radiation	4	4th- C2.3		PS4, PS8	46.67	Low MOSART
PS	Information Technologies and Instrumentation	4					New to OAS
PS	Chemical Reactions	5			PS5	33.3	Low MOSART
PS	Types of Interactions	5					New to OAS

Note: PASS = Priority Academic Student Skills; OCCT = Oklahoma Core Curriculum Tests (OCCT); STEBI-A = Science Teaching Efficacy Beliefs Inventory–Form A; MOSART = Misconception-Oriented Standards-Based Assessment Resources for Teachers; ESS = Earth and Space Science; LS = Life Science; PS = Physical Science

Appendix F: EDSTEMA[1] Science Content Knowledge (SCK) Assessment

	Question	Origin
1	By examining fossils, scientists think that: a. all existing plants and animals have been unchanged since Earth formed. b. plants have changed over time, but animals have not. c. animals have changed over time, but plants have not. d. many changes have occurred to plants and animals. e. It is not possible to determine the past from fossils.	MOSART[2] 5–8 Earth Science Test Form 821 #1
2	Below is a diagram of Earth's interior. What is layer 2? a. Crust b. Mantle c. Outer core d. Inner core e. Lithosphere 	MOSART 5–8 Earth Science Test Form 821 #2
3	The flu virus has most likely spread because: a. humans are slowly being exterminated. b. it keeps adapting to new environments. c. flu virus wants to infect people everywhere. d. the virus is smarter, faster and stronger than most people. e. overcrowding and pollution keep it alive and contagious.	MOSART 5–8 Life Science Test Form 921 #16
4	In the picture below, which parts make up a desert ecosystem?  a. Only the cactus. b. Only the sand. c. Only the cactus and the sand. d. The cactus, sand and sunlight. e. A desert is not an ecosystem.	MOSART 5–8 Life Science Test Form 921 #20

[1] Every Day STEM for All
[2] Misconception-Oriented Standards-Based Assessment Resources for Teachers

5	Blindness can be caused by damage to which of the following? a. Only the eyes. b. Only the brain. c. Only the spinal cord. d. Either the eyes or the brain. e. Either the brain or the spinal cord.	MOSART 5–8 Life Science Test Form 921 #21
6	In a forest ecosystem foxes feed mostly on a variety of birds and rodents (small furry mammals). What would probably happen to the foxes if a flood destroyed one type of rodent? The foxes would: a. eat other rodents and birds. b. eat berries. c. eat insects. d. die. e. move elsewhere.	MOSART 5–8 Life Science Test Form 921 #22
7	Which of the following are composed of cells working together? a. Only blood and muscle. b. Only blood and bone. c. Only muscle and bone. d. Blood, muscle and bone. e. None of the above.	MOSART 5–8 Life Science Test Form 921 #23
8	A light bulb is connected to a battery by wires. The bulb is lit up. Nadia wants to know what is flowing through the wires. If a scientist were to cut the wire and look at it with a powerful magnifying glass, what do you think she would see?  a. Chemicals from the battery flowing through the wire. b. Light flowing through the wire. c. The wire will be hollow with nothing flowing through it. d. Tiny sparks flowing through the wire. e. The wire will be solid.	MOSART 5–8 Physical Science Test Form 721 #4
9	Imagine that you go to leave a room with an overhead light. The light is on. You move the wall switch to turn off the light, but the light stays on. What is probably wrong? a. The battery that powers the switch is dead. b. There is a break in the wire to the light bulb. c. The switch can no longer stop the flow of electricity through the wires d. The light fixture is broken. e. A surge of electricity is occurring in the building.	MOSART 5–8 Physical Science Test Form 721 #18

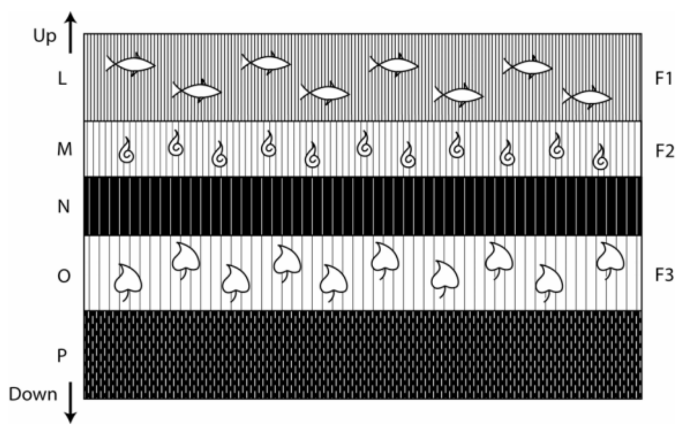
10 If the temperature of Earth's interior was the same as at the surface, then the movement of the continents would: — MOSART 9–12 Earth Science Test Form 831 #3

a. stop.
b. slow down.
c. stay the same.
d. speed up.
e. There is no way to predict how motion of the continents would be affected.

11 One type of fish fossil is found only in a rock layer near the bottom of a canyon formed in undisturbed rock. A different kind of fish fossil is found only in a rock layer 4,000 feet higher, near the top of the canyon. Which statement would a scientist most likely offer to explain this situation? — MOSART 9–12 Earth Science Test Form 831 #8

a. The fish in the lower layer lived in deeper water than the other fish.
b. The fish in the upper layer lived more recently than the other fish.
c. The rock in the lower layer is denser than the rock in the upper layer.
d. The rock in the upper layer must have formed in fresh water.
e. None of the above; there is not enough information provided.

12 A scientist examines a series of undeformed rock layers (see figure below). The most straightforward choice for the oldest fossil is: — MOSART 9–12 Earth Science Test Form 831 #13



a. F1
b. F2
c. F3
d. they are all the same age.
e. An estimate cannot be made from the diagram.

13 Which of the following is the most similar to mantle convection? — MOSART 9–12 Earth Science Test Form 831 #14

a. Smoke rising above a fire on a calm day.
b. Ocean water being stirred up by surface winds.
c. Lava collecting into pools at the base of a volcano.
d. Mud sinking to the bottom of a lake.
e. Sand moving as dunes across the desert.

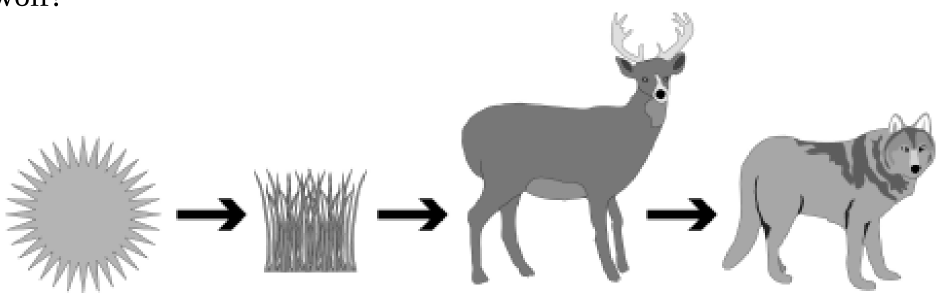
14 Convection of Earth's mantle will cause Earth's:
a. spin to reverse.
b. interior to cool down.
c. surface to be covered with lava.
d. oceans to boil away.
e. axis to change its tilt

MOSART 9–12 Earth Science Test Form 831 #18

15 Copper is a good electrical conductor because:
a. it has a high melting temperature.
b. no electrons can flow through it.
c. few electrons can flow through it.
d. electrons flow readily through it.
e. it can also be a good insulator

MOSART 9–12 Physics Test #11

16 Given the food chain shown below, what energy would be available to the wolf?



a. All of the energy from the Sun, grass and deer.
b. All of the energy from the Sun and grass, minus the energy in the deer.
c. All of the energy transferred from the Sun to the grass.
d. Only the energy contained in the grass.
e. Some of the energy transferred from the Sun to the grass to the deer.

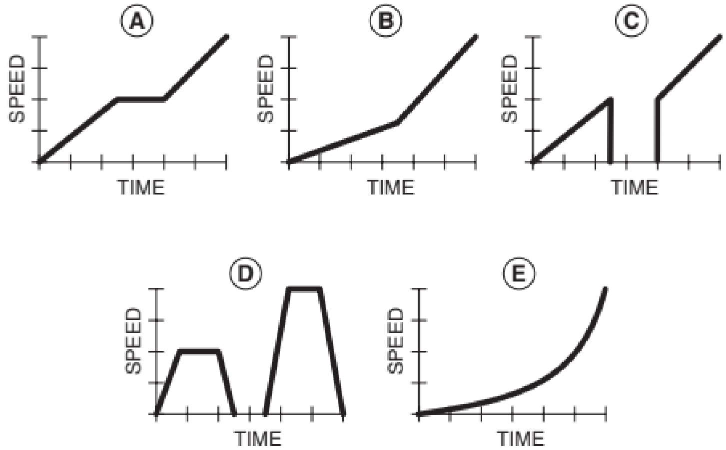
MOSART 5–8 Life Science Test Form 921 #23

17 A population of cave-dwelling fish used to include individual fish with functioning eyes, but now none of the fish have functioning eyes. The most scientifically accurate explanation for this is that:
a. the fish decided that they no longer needed eyes and stopped developing them.
b. fish with functioning eyes were less likely to survive to reproduce.
c. individual fish adapted to living in the cave during their lifetime and no longer needed eyes.
d. cave conditions produced mutations that caused the fish to lose their eyes.
e. the strongest fish survived, which happened to be the ones without eyes.

MOSART Life Science Test for Educators #30

18 Carolyn walks a half mile to school. One morning, halfway to school, she stopped to watch a bird building a nest. When she realized she was late, she ran the rest of the way to school to avoid being marked late. Which graph below shows Carolyn's speed during her walk to school?

MOSART 5–8 Physical Science Test Form 721 #7



19 After a light wave has reflected from a smooth glass mirror hanging on a wall:
   a. It may be traveling in a different direction.
   b. It is traveling at a different speed than before it hit the mirror.
   c. Its wavelength is different.
   d. Its energy is greater.
   e. All of the above.

MOSART 9–12 Physics Test #25

20 Birds that live on some islands get their food by cracking nuts with their beaks. Which birds are most likely to survive on an island where nuts are the main type of food?
   a. Birds with soft beaks.
   b. Birds with no beaks.
   c. Birds with strong beaks.
   d. Birds with dark-colored beaks.
   e. Birds with light-colored beaks.

MOSART Life Science Test for Educators #8

Appendix G: Item Responses- Science Content Knowledge (SCK) Assessment

Item #	Treatment Group Pretest % Correct	Treatment Group Posttest % Correct	Treatment Group Pre-Post Change	Comparison Group Pretest % Correct	Comparison Group Posttest % Correct	Comparison Group Pre–Post Change
1	100	95	-5	94	95	1
2	73	82	9	53	61	8
3	89	93	4	76	86	10
4	30	50	20	24	25	1
5	77	91	14	71	75	4
6	80	70	-10	71	66	-15
7	77	77	0	65	66	1
8	55	57	2	41	43	2
9	73	91	18	65	73	8
10	9	55	46	18	18	0
11	73	89	16	65	70	5
12	91	98	7	88	91	3
13	9	11	2	12	11	-1
14	23	20	-3	24	5	-19
15	80	100	20	71	80	9
16	75	82	7	65	64	-1
17	20	30	10	18	16	-2
18	80	82	2	76	75	-1
19	75	59	-16	53	52	-1
20	61	75	14	59	57	-2
Average	59	75	16	55	54	-1

Appendix H: Descriptive Statistics—Science Content Knowledge (SCK) Assessment

	Treatment Group			Comparison Group		
	Pretest	Posttest	Difference	Pretest	Posttest	Difference
Number	44	44	44	17	17	17
Minimum	0.2	0.45	-0.05	0.15	0.25	-0.2
Maximum	0.85	0.95	0.45	0.8	0.75	0.15
Range	0.65	0.5	0.5	0.65	0.5	0.35
Mean	0.59	0.75	0.16	0.55	0.51	-0.01
Median	0.6	0.75	0.15	0.6	0.6	0
Mode	0.6	0.7	0.15	0.6	0.6	0
Standard deviation	0.15	0.11	0.10	0.19	0.16	0.11

Appendix I: Mean Scores—STEBI-A

	Treatment			Comparison		
Statement	Pre	Post	Diff	Pre	Post	Diff
1. When a student does better than usual in science, it is often because the teacher exerted a little extra effort.	3.0	3.7	0.7	3.1	3.7	0.6
2. I am continually finding better ways to teach science	4.0	4.4	0.4	4.2	4.3	0.1
3. Even when I try very hard, I don't teach science as well as I do most subjects.	3.2	3.5	0.3	2.8	2.8	0
4. When the science grades of students improve, it is most often due to their teacher having found a more effective teaching approach.	3.8	3.8	0	3.7	4.0	0.3
5. I know the steps necessary to teach science concepts effectively.	3.4	4.0	0.6	3.1	3.4	0.3
6. I am not very effective in monitoring science experiments.	3.4	3.8	0.4	3.6	3.7	0.1
7. If students are underachieving in science, it is most likely due to ineffective science teaching.	3.1	3.7	0.6	3.1	3.0	-0.1
8. I generally teach science ineffectively.	3.7	3.9	0.2	3.7	3.8	0.1
9. The inadequacy of a student's science background can be overcome by good teaching.	3.7	3.9	0.2	3.2	3.1	-0.1
10. The low science achievement of some students cannot generally be blamed on their teachers.	3.0	3.3	0.3	3.4	3.3	-0.1
11. When a low-achieving child progresses in science, it is usually due to the extra attention given by the teacher.	3.4	3.8	0.4	3.8	3.7	-0.1
12. I understand science concepts well enough to be effective in teaching elementary science.	3.8	4.3	0.5	3.5	3.8	0.3
13. Increased effort in science teaching produces little change in some students' science achievement.	3.6	3.6	0	3.6	3.5	-0.1
14. The teacher is generally responsible for the achievement of students in science.	3.4	3.8	0.4	3.4	3.4	0
15. Students' achievement in science is directly related to their teacher's effectiveness in science teaching.	3.3	3.8	0.5	3.4	3.5	0.1
16. If parents comment that their child is showing more interest in science at school, it is probably due to the performance of the child's teacher.	3.5	3.8	0.3	3.6	3.5	-0.1
17. I find it difficult to explain to students why science experiments work.	3.6	3.9	0.3	3.5	3.8	0.3
18. I am typically able to answer students' science questions.	3.7	4.0	0.3	3.6	3.8	0.2
19. I wonder if I have the necessary skills to teach science.	3.3	3.9	0.6	3.4	3.5	0.1
20. Effectiveness in science teaching has little influence on the achievement of students with low motivation.	3.7	3.9	0.2	3.7	3.7	0

21. Given a choice, I would not invite the principal to evaluate my science teaching.	3.9	4.2	0.3	3.8	3.6	-0.2
22. When a student has difficulty understanding a science concept, I am usually at a loss as to how to help the student understand it better.	3.7	3.9	0.2	3.7	3.8	0.1
23. When teaching science, I usually welcome student questions.	4.1	4.4	0.3	4.1	4.3	0.2
24. I don't know what to do to turn students on to science.	3.7	3.9	0.2	3.7	3.7	0
25. Even teachers with good science teaching abilities cannot help some kids learn science.	3.3	3.5	0.2	3.3	3.3	0
Total	88.4	95.8	7.4	90.1	90.0	-0.1

Note: Response scale: 1 = *strongly disagree*, 2 = *disagree*, 3 = *uncertain*, 4 = *agree*, 5 = *strongly agree*. STEBI-A = Science Teaching Efficacy Beliefs Instrument–Form A.

Appendix J: STEBI-A Construct Scores

Construct	Treatment			Comparison		
	Pre	Post	Diff	Pre	Post	Diff
Outcome Expectancy (Items 1, 4, 7, 9–11, 13–16, 20, 25)	3.4	3.72	0.32	3.44	3.48	0.0
Personal Teaching Efficacy (Items 2, 3, 5, 6, 8, 12, 17–19, 21–24)	3.65	4.01	0.35	3.59	3.72	0.12

Note: STEBI-A = Science Teaching Efficacy Beliefs Instrument–Form A; Pre = pretest; Post = posttest; Diff = posttest minus pretest.

Appendix K: Descriptive Statistics–- STEBI-A

	Treatment Group		Comparison Group	
	Pretest	Posttest	Pretest	Posttest
Number	40	40	17	17
Minimum	69	80	70	75
Maximum	106	109	106	102
Range	37	29	36	27
Mean	88.3	96.0	87.7	88.3
Median	89	97	88	88
Mode	89	98	93	80
Standard deviation	8.50	8.75	9.27	7.89

Note: STEBI-A = Science Teaching Efficacy Beliefs Instrument–Form A

Appendix L: Descriptive Statistics–- OCCT Scores

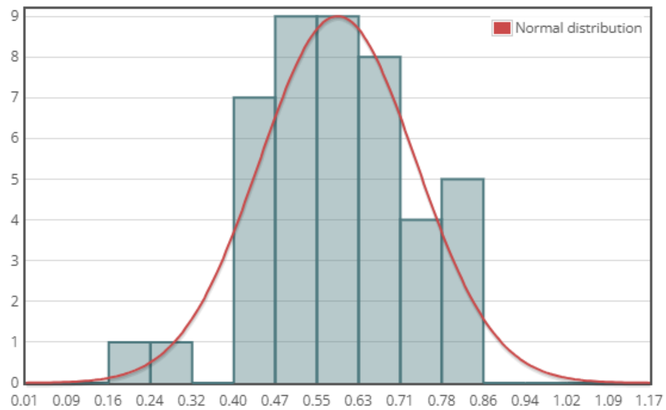
	Year 0	Year 1
Number	27	27
Minimum	21	35
Maximum	100	100
Range	79	65
Mean	70.15	79.41
Median	74	79
Mode	55	94
Standard deviation	18.777	15.797

Note: OCCT = Oklahoma Core Curriculum Test.

Appendix M: Figures

Figure M1

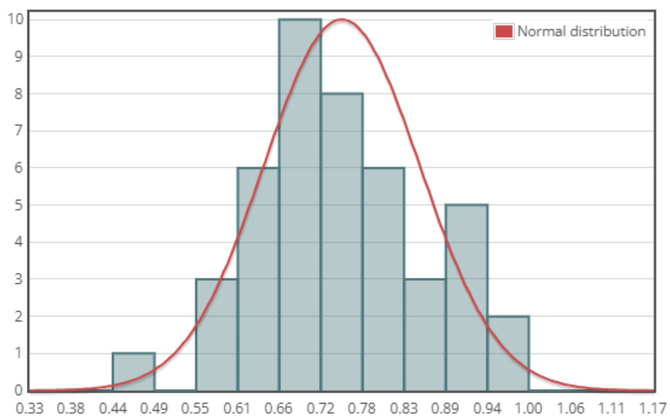
Normal distribution of treatment group Science Content Knowledge pretest scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222152527919

Figure M2

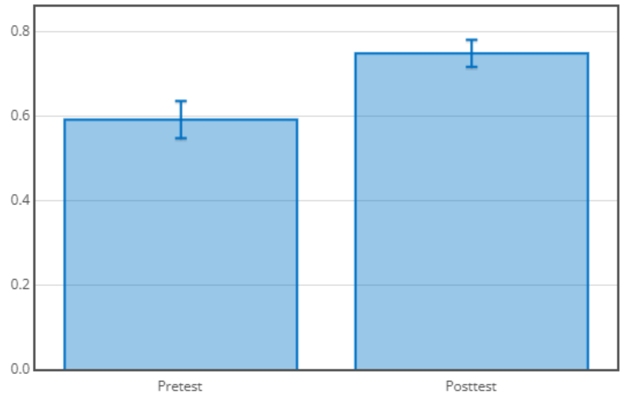
Normal distribution of treatment group Science Content Knowledge posttest scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222153222088

Figure M3

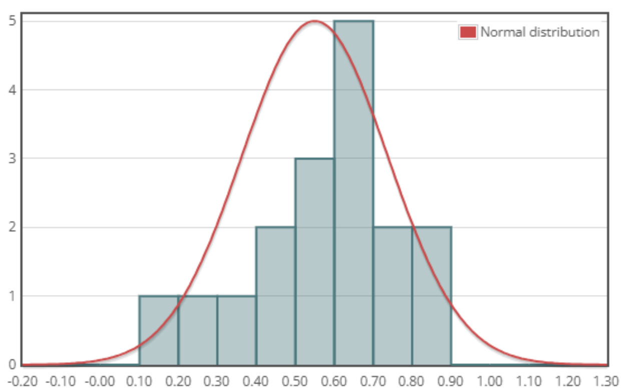
Pretest and posttest scores earned by treatment group on the Science Content Knowledge Assessment



Note: The error bars on the graph are 95% confidence intervals. From *AICBT (2025).* AI–Therapy Statistics *(BETA version)* [Comparing two sets of data]. https://www.ai-therapy.com/psychology-statistics/results/20250222181103573

Figure M4

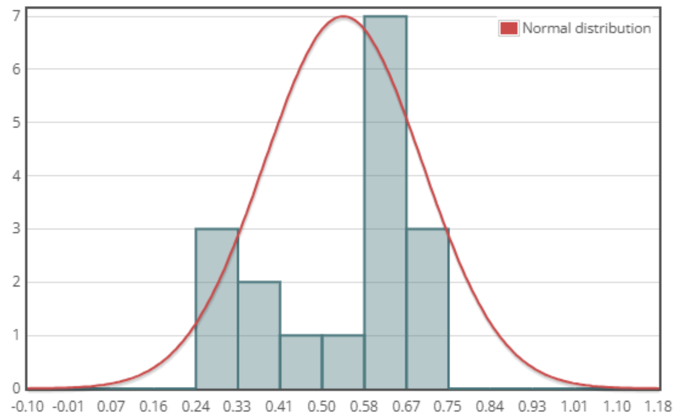
Normal distribution of comparison group pretest scores on the Science Content Knowledge Assessment



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222153650088

Figure M5

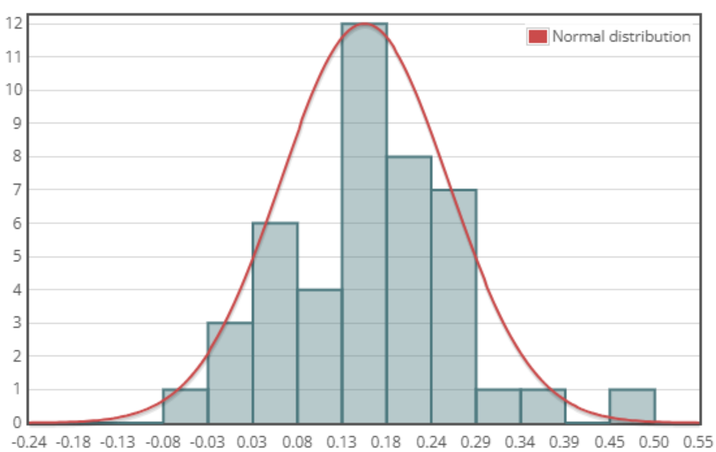
Normal distribution of comparison group posttest scores on the Science Content Knowledge Assessment



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222154700205

Figure M6

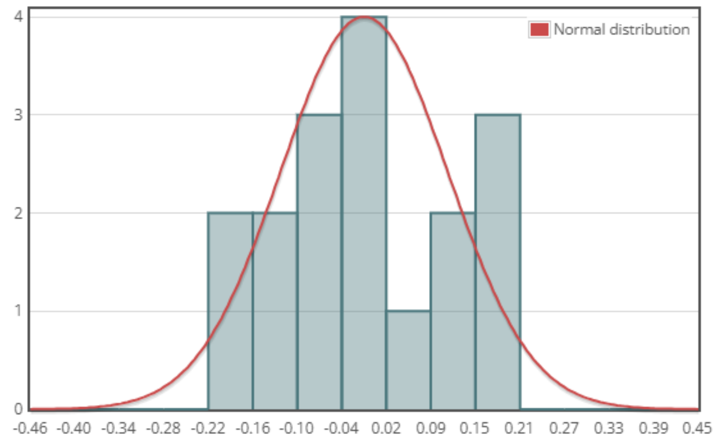
Normal distribution in change in treatment group Science Content Knowledge Assessment scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222155339468

Figure M7

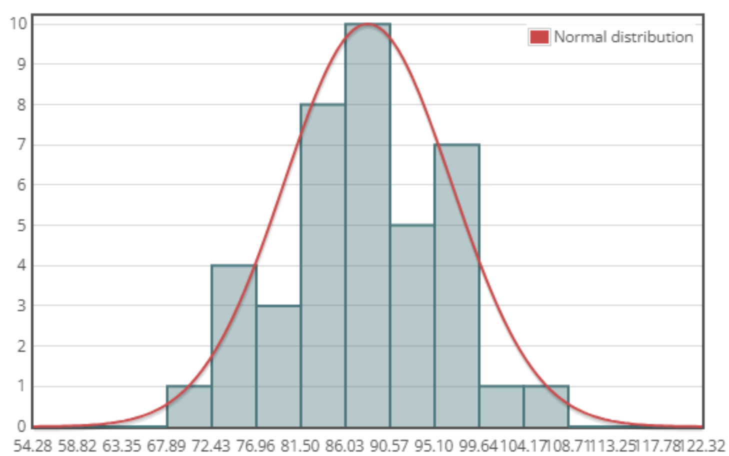
Normal distribution in change in comparison group Science Content Knowledge Assessment scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222160450443

Figure M8

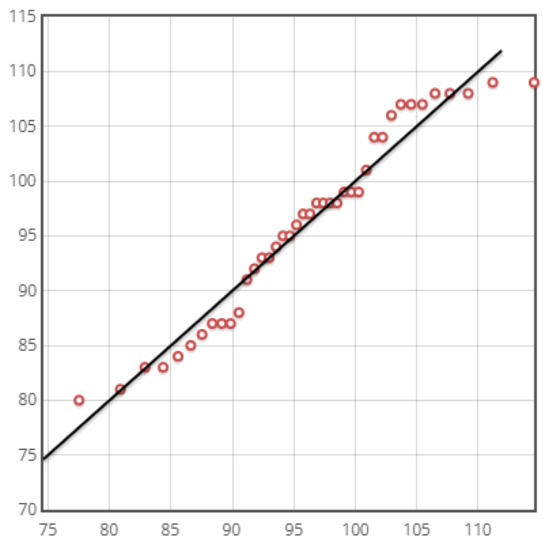
Normal distribution of treatment group pretest scores on STEBI-A



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222160948727

Figure M9

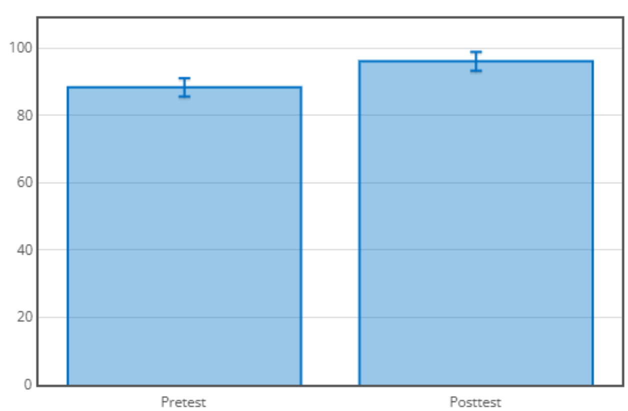
Q–Q plot for treatment group STEBI-A posttest scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222180412355

Figure M10

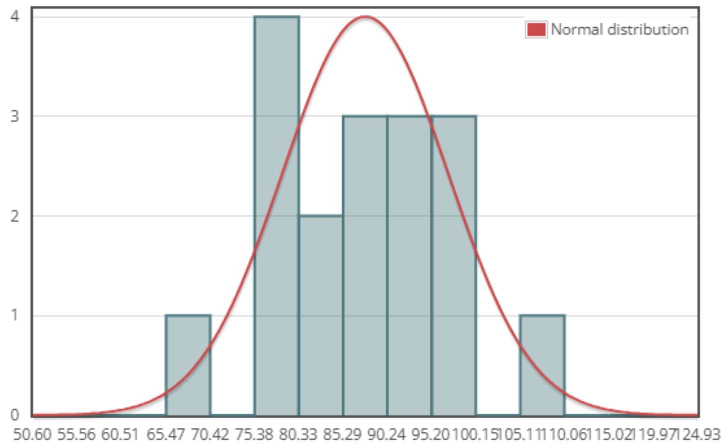
Pretest and posttest scores earned by treatment group on the STEBI-A



Note: The error bars on the graph are 95% confidence intervals. From *AICBT (2025).* AI–Therapy Statistics *(BETA version)* [Comparing two sets of data]. https://www.ai-therapy.com/psychology-statistics/results/20250222185040563

Figure M11

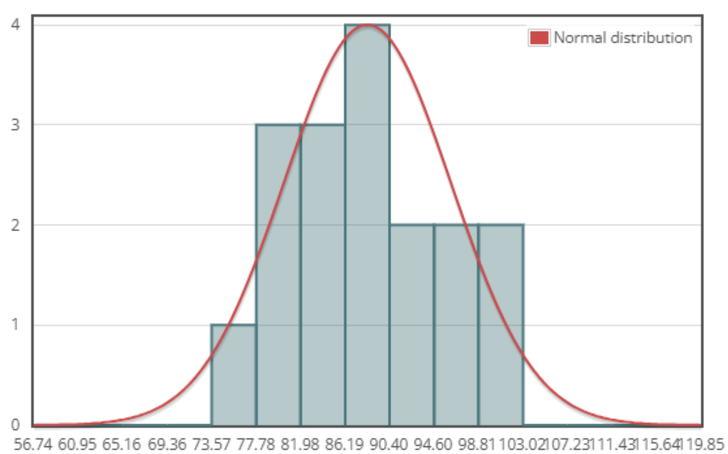
Normal distribution of comparison group STEBI-A pretest scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222193423819

Figure M12

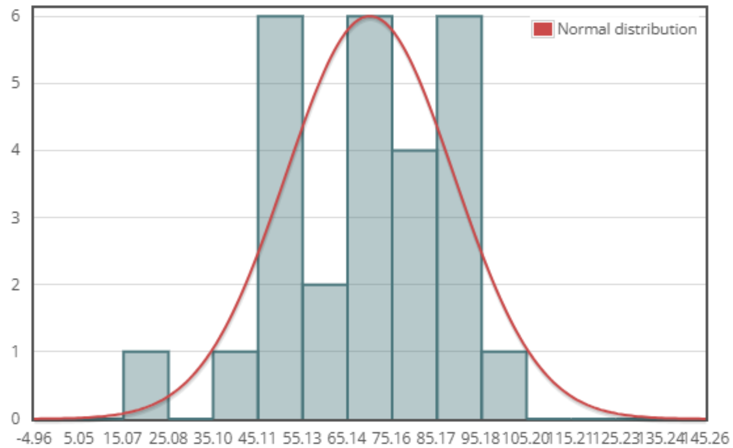
Normal distribution of comparison group STEBI-A posttest scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222191737246

Figure M13

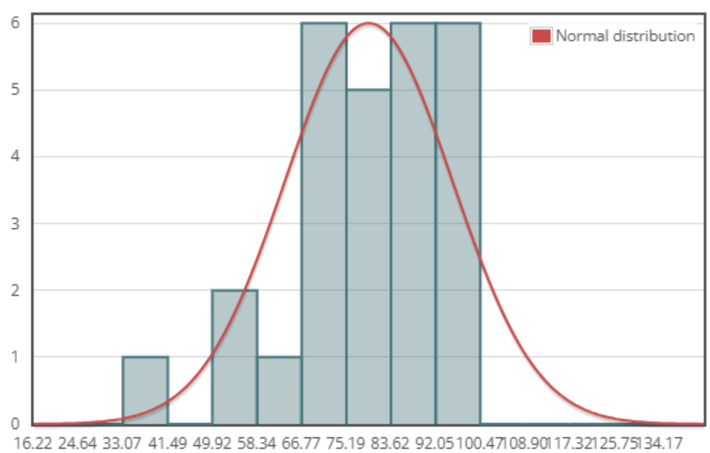
Normal distribution of Year 0 fifth-grade science Oklahoma Core Curriculum Tests scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/20250222211502134

Figure M14

Normal distribution of Year 1 fifth-grade science Oklahoma Core Curriculum Tests scores



Note: From AICBT (2025). AI–Therapy Statistics *(BETA version)* [The normal distribution]. https://www.ai-therapy.com/psychology-statistics/results/2025022221