NMSI Middle School Longitudinal Study

Final Report

Dr. Richard Brown

Submitted to: National Math + Science Initiative

Date: October 20, 2024

Laying the Foundation Teacher Training and Student Outcomes

Executive Summary

The National Math and Science Initiative (NMSI) aims to enhance teaching quality in Science, Technology, Engineering, and Mathematics (STEM) fields. This report evaluates the impact of its programming, focusing on the effects of its Laying the Foundation programming on both middle school teachers and their students in various locations over the course of three years (2021-2024). The program has been restructured to provide additional resources, tools, and ongoing support throughout the academic year, aiming to improve teacher effectiveness and student outcomes, especially in underrepresented groups.

The primary focus of the evaluation centers on two research questions:

1. Impact on Student Mindsets: How does the NMSI program influence student interest in STEM, self-identification with STEM subjects, and STEM career aspirations?
2. Impact on Student Achievement: How does the NMSI program affect student academic performance in algebraic and statistical reasoning, and how do these effects vary across different student subgroups?

Key Findings

Student Mindset Improvements

The report shows that students taught by NMSI-trained teachers demonstrated a significant improvement in their confidence and interest in STEM subjects. Students' perceptions of their ability to succeed in STEM courses and interest in pursuing STEM careers increased, with particularly strong gains among Black students who showed a notable rise in STEM career aspirations. The program did not result in a uniform increase in career aspirations across all subgroups, with Hispanic students showing relatively lower growth in STEM career interest.

Academic Performance Gains

There were measurable improvements in algebraic and statistical reasoning among students taught by NMSI-trained teachers. Gains were particularly pronounced in algebraic reasoning, with students improving their problem-solving skills significantly over the academic year. The improvements in academic performance were consistent across gender and ethnic groups, indicating the program’s effectiveness in supporting a diverse range of students.

Inclusivity and Impact on Underrepresented Students

The program was particularly effective in supporting underrepresented groups, including female, Black, and Hispanic students, who demonstrated academic gains similar to their peers. The study

underscores the potential of the NMSI program to close achievement gaps by fostering an inclusive learning environment, where students from all backgrounds are given equal opportunities to excel in STEM fields.

Program Design and Methodology

The NMSI program is designed around a three-year progression for teacher training, which includes summer training followed by ongoing support throughout the school year. The training emphasizes inquiry-based learning: Encouraging critical thinking and problem-solving, instructional scaffolding, providing differentiated instruction tailored to varying student needs, and intentional course progression by aligning teaching strategies with state standards to ensure coherent learning paths across grade levels. Data was collected through student assessments and surveys focused on STEM perceptions and algebraic/statistical reasoning skills. The study evaluated the impact of the NMSI program using these measures across three cohorts of teachers and students in different academic years (2021-2024).

Conclusions and Recommendations

The NMSI program has proven to be an effective professional development initiative that not only enhances teaching practices but also drives measurable improvements in student outcomes. The key strengths of the program include:

- Significant improvements in student confidence and interest in STEM subjects.
- Measurable academic gains in algebraic and statistical thinking.
- Positive outcomes across diverse student groups, helping to close achievement gaps in STEM education.

The report suggests that the success of the NMSI program could serve as a model for future initiatives aimed at improving STEM education. Investing in teacher professional development, as demonstrated by the NMSI program, can play a critical role in fostering student success and preparing the next generation for careers in STEM fields. As educational institutions and policymakers seek to improve STEM education, the findings from this project provide valuable insights into the effectiveness of content-rich, sustained professional development programs like NMSI provides.

Strategic Implications

The NMSI program's results highlight the importance of continued investment in teacher development to improve instructional practices in STEM and emphasizing inclusivity by addressing the unique needs of underrepresented student populations, which can lead to more equitable academic outcomes. This executive summary showcases the significant potential of well-structured professional development programs in transforming both teacher practices and student achievements in STEM education.

Laying the Foundation Teacher Training and Student Outcomes

Abstract

This study evaluates the impact of the enhanced "Laying the Foundation" (LTF) professional development program by the National Math and Science Initiative (NMSI) on middle school teachers and their students across three different locations. NMSI, a leader in educational reform, refined its well-established program by introducing additional resources, tools, and ongoing support aimed at improving teaching practices, particularly in STEM (Science, Technology, Engineering, Mathematics) education. The study focused on two primary research questions: (1) How does the NMSI program influence student mindsets, including their interest in STEM courses, self-identification with STEM subjects, and STEM career aspirations? (2) How does this NMSI program impact student achievement, specifically in algebraic and statistical thinking, and how do these effects vary across different student subgroups?

The research involved three cohorts of teachers and students over three academic years (2021-2024). Data were collected through assessments measuring student perceptions of STEM and their achievement in algebraic and statistical reasoning. The findings indicate that students taught by NMSI-trained teachers exhibited significant improvements in their confidence in STEM subjects, with measurable gains in both algebraic and statistical thinking. These improvements were consistent across all cohorts and were particularly pronounced in algebraic reasoning.

Notably, the study found that Black students showed increased interest in STEM careers, while Hispanic students demonstrated less increase in this area compared to their peers. However, academic achievement gains were consistent across gender and ethnic boundaries, underscoring the program's effectiveness in supporting diverse student populations.

This study highlights the positive impact of sustained, content-rich professional development on both teacher effectiveness and student outcomes. The NMSI program's emphasis on inquiry-based learning, instructional scaffolding, and intentional course progression contributed to significant gains in student confidence and academic performance in STEM subjects. The findings suggest that the NMSI program effectively helps close achievement gaps and fosters an inclusive environment where all students have the opportunity to excel in STEM, offering valuable insights for future educational initiatives aimed at enhancing STEM education and teacher development.

Laying the Foundation Teacher Training and Student Outcomes

Introduction

This project examines the effects of an enhanced professional development experience of middle school teachers in dozens of classrooms using student perceptions and student learning outcomes. During this project, the National Math and Science Initiative (NMSI) refined the resources, tools, and training experience of its well-known, in-demand Laying the Foundation (LTF) summer professional development experience and expanded supports into the school year.

For nearly 30 years, teachers and students have benefitted from training, support and resources provided by the National Math and Science Initiative (NMSI), formerly known as AP Strategies. NMSI partners with schools and districts nationwide to provide extraordinary training and support for teachers and to give all students the resources they need to develop and demonstrate knowledge and skills that will propel them throughout their lives. To date, NMSI's programs have impacted 2 million students, 70,000+ teachers,1,300+ high schools, and 44 universities.

All NMSI programs are guided by a shared core belief: Unlocking student potential starts with great teachers. NMSI’s programs are well-documented to increase academic intensity and access to rigorous courses, improve student achievement, and decrease the **college readiness gap,**[1] especially among underrepresented students. Funded by multiple, significant grants from the US Department of Education and the Department of Defense, NMSI’s College Readiness Program’s (CRP) has been studied in various settings, producing a growing body of evidence indicates that CRP not only increases the effectiveness of teachers as measured by raising the probability that students will take and earn qualifying scores on AP exams, hence increasing their achievement and college readiness, but also has significant and longer term positive postsecondary and economic impacts. The program’s consistent elements produce reliably successful and sustained outcomes across settings, states, subject areas, teachers, and students, including in schools with students traditionally underrepresented in AP courses. For example, in 2017-18, NMSI worked with 19 Texas districts through CRP focused on the AP CSP course, training 21 AP CSP teachers. This work resulted in a 330% Year 1 year-over-year enrollment increase, and a 327% increase in year-over-year qualifying scores. NMSI’s impact on AP access and qualifying scores in general, as summarized in Figure 3, points to promising results, particularly for underrepresented students.

Similar to CRP, NMSI’s Laying the Foundation gives teachers the resources they need to raise expectations and develop advanced levels of thinking and learning. The program aims to train teachers to facilitate students’ progression through the academic pipeline toward college-

[1] For purposes of NMSI’s College Readiness Program, the college readiness gap is measured by the number of high-need students who take and earn qualifying scores on AP exams because the AP exam is one of the few nationally accepted proxies for college readiness.

level coursework through a three-year teacher training progression; each year's four-day summer institute provides hands-on training, classroom-ready materials, and instructional best practices. Trainers guide teachers through content-rich instruction that moves beyond what to teach, to how to deepen student understanding of key concepts. The program also offers classroom-ready materials and resources which are aligned with state standards and encourage higher-order thinking.

NMSI training emphasizes research-based instructional strategies including: inquiry-based learning, instructional scaffolding (techniques and guidance for delivering differentiated instruction), and intentional course progression (education about the knowledge and skills that students need to master at each grade level) all designed to help increase academic rigor and build college and career readiness (Friedrichsen & Berry, 2015). The training is focused on the teacher experience, allowing teachers to enact practices in a safe professional development setting before using these instructional strategies in their classroom. Teachers gain these strategies/tools through a practice-based, three-year learning progresssion (Schneider & Plasman, 2011):

- Year One: Thinking about Learners, where teachers learn and practice delivering high-quality lessons while gathering feedback from other educators.
- Year Two: Thinking about Teaching, where teachers analyze and reflect on how they prepare and lead classes and learn best practices from expert teachers and other educators; and
- Year Three: Choosing a Pathway, where teachers select the development path that suits their individual needs.

There is one research study on the efficacy of LTF. Published in 2017 (Brown & Phelan, 2017), it measured the impact and success of LTF on student academic success in a matched set of treatment and control schools in Alabama, using student level ACT/Aspire data for the 2015-1016 academic year as the primary outcome variable to measure change in student achievement. Significant effects were found for math in both grades 8 and 10. For 8th grade math, treatment students scored on average 3.217 points higher on the ACT/Aspire exam than control students while holding constant prior math achievement (Cohen's D effect size of .47). Grade 10 analyses showed similar results, though the effects are slightly smaller. Taken together, these results provide evidence that students of LTF-trained teachers outperformed students with comparable academic achievement of teachers not receiving the LTF training across grade level and subject area.

Theoretical Foundation

Historically, professional development initiatives have had mixed results when considering student achievement outcomes as indicators of success (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007; Fischer, et al., 2018). Possible reasons for this include poor implementation of professional development programs, as well as lack of adherence to characteristics of professional development programs shown to have positive effects. These characteristics include a focus on pedagogical content knowledge and skills of the teachers, ongoing and collaborative nature, as well as an emphasis on teacher practices (Garet, Porter, Desimone, Birman, Herman & Yoon, 1999; Hiebert, 1999). Most commonly reported in the literature are professional development programs which are short, "one-shot" workshops. For example, Birman et al., (2007) reported that few teachers receive intensive, sustained, and content-focused professional development in mathematics. Their data indicated that while teachers report taking part in professional development activities focused on teaching—very few took part in these activities for an extended period; teachers averaged 8.3 hours of professional development on how to teach mathematics and 5.2 hours of more in-depth study. In contrast, NMSI teachers receive 27-30 hours of training per year.

Significant Duration and Implementation-Driven Experiences. NMSI's program builds and strengthens teachers' skills over a three-year period, rooted in the core belief that all teachers can put all students on track for AP success if they have a deep understanding of content knowledge, pedagogic content knowledge (PCK), instructional planning, data analysis, assessment, and intentional course alignment. In most of our content areas, teachers are introduced to major concepts in Year One and then dive deeper in the nuances of the concepts and have more opportunities to practice teaching those concepts at an advanced level in Years Two and Three. An additional focus is building on implementation-driven learning experiences that over time allow for teachers to illustrate their own practice and share feedback during the trainings.

Materials are designed to grow teacher skills in assessment, data analysis, and intentional course progression, in addition to PCK and instructional planning, while providing a cohesive experience from one day to the next. This cohesion may be achieved through the use of unit-based materials. This design principle is in response to the twin desires to keep content at the center while transforming teacher practice. By designing the sessions with an eye to both skill-based and content objectives while being cautious of redundancy, experiences can be crafted that are high-leverage, novel, and high-impact. Teachers leave every programming event with materials ready for adoption and adaption and are provided with opportunities to rehearse these during the sessions. Such materials also provide exemplars that help teachers form "mental representations" of excellence. This program offers:

- Adaptable lesson plans and units/materials, which are the cornerstone of NMSI trainings. The program supports teacher content knowledge and pedagogy through excellent materials that ultimately align to college-level coursework.
- Support in modifying materials for the various contexts in which teacher work, including a focus on scaffolding and enrichment to ensure that learning is always a "productive struggle" (Hammond, 2008).

To positively impact teacher practices, opportunities must be provided for teachers to enact those practices and receive feedback during the program. When possible, the practicing of practice should include the following elements: (1) inquiry-driven explorations and deconstruction of the practice based on lesson artifacts; (2) enacting of the practice; (3) feedback on the practices; (4) reflection on and plan to implement the practice. During trainings, teachers are provided with a mental representation of great teaching of a given concept by allowing them to immersively experience a model, to engage in pedagogical discussions on what makes the model effective, and to then work on their own classroom materials through the modification of lessons, the making of plans, and role-playing the instruction they would facilitate (Ball & Cohen, 1999). Learning from teacher practices will be enhanced by the inclusion of classroom artifacts beyond the lesson plan that allow for a deeper understanding of the teacher practices and how to enact them. Classroom artifacts will form the basis of teachers' mental representation of what practice can look like; for example:

- Video recording of lessons and/or activities enacted in a real classroom
- Samples of student work and/or modified lesson materials
- Teacher narratives/reflections

Figure 1. Levels of Implementation

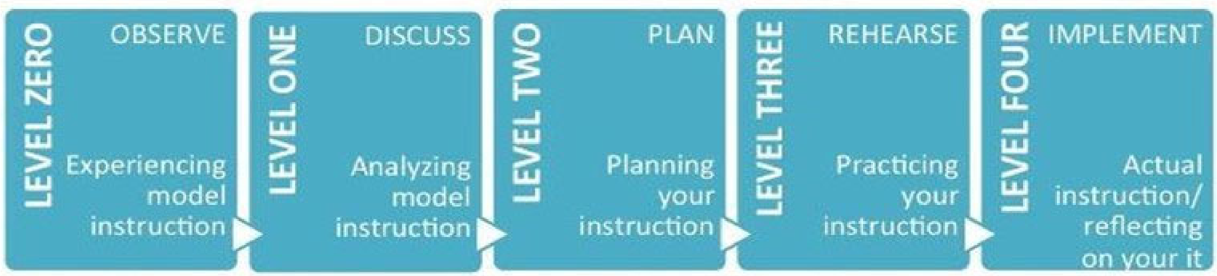

Figure 1 represents the progression of the NMSI professional development as a sequence of five levels; all should be present in the program experience; Levels Zero through Three are present throughout the experience, with Level Four happening largely during the academic year when teachers are applying their learning in their own classrooms. Teachers also had access to instructional coaches and online discussion boards during the school year to ask questions, collaborate, and share resources.

Intentional course progression will allow for a coherent student experience. A key component of this NMSI program is the establishment of intentional course progression, which helps ensure that students' experiences across grade levels are coherent and that what they learn during one grade builds to the next or from the last. Building in structures for intentional course progression will allow teachers across grade levels to come together to plan and tailor instruction. Participating teachers are introduced to intentional course progression to identify and examine any persistent challenges preventing underrepresented students from participating in the pathway to college-level courses, and to observe mindsets and attitudes that could be obstacles to students persisting in the pathway. For example, opportunities are provided to explore what the college-level version of a given skill looks like, not to teach to the test but to understand what the full development of skill looks like, with time in every day of training to work with materials for the grade level at which a teacher works. By design, if a district had teachers from every grade participate over the course of the program, their students would be ready for college-level coursework by 11th grade.

Culturally responsive teaching. NMSI's definition of Culturally Responsive Teaching echoes the work of Geneva Gay (2010) and Zaretta Hammond (2015). Culturally responsive teaching is using students' "cultural knowledge, prior experiences, frames of reference, and performance styles...to make learning encounters more relevant to and effective for them" (Gay, p. 31). In alignment with the research on culturally responsive teaching practices, NMSI conceives of "culture" as the lens through which each individual views the world, a lens impacted by the intersectionality of their identity components (including, but not limited to race, ethnicity, religion, socio-economic class, gender, and sexuality) and their lived experiences. All students are complex and knowledgeable, carrying with them meaningful past experiences that should inform their classroom experiences. Hammond (2015) claims that "culturally responsive teaching is a pedagogical approach firmly rooted in learning theory and cognitive science" (p. 16) and her focus on relationships among students, teachers, and communities as cornerstones in NMSI's work. NMSI maintains that teachers and educators must aim to provide all students with the scaffolds that "promote effective information processing" and build students' "intellective capacity," ensuring that they become increasingly independent learners with strong higher order thinking skills (Hammond, p. 16).

NMSI's intention in LTF is to positively impact students who are furthest from opportunity, therefore it is essential to recognize the ways in which school systems can be limiting for students who are not middle-class and white and work to create environments and classrooms that see students' differences as strengths and leverage them accordingly. When providing practice opportunities, NMSI encourages teachers to think about the cultures of their students and the strengths they may bring into the classroom, considering how these strengths can be leveraged within lessons. NMSI has curated culturally responsive protocols that guide teachers in making modifications for their students, providing rationales for when and why to use the different protocols. They also provide learning experiences for teachers that name and explicitly explore what it means to be increasingly culturally responsive.

Rationale and Importance

Delivered to more than 50,000 teachers in 1,800 districts across the country over the past 10 years, LTF has been well-received, creating an AP course pipeline for more than one million students (NMSI Internal Data, 2019). Recently, NMSI saw an opportunity to enhance the three-year progression for increased efficacy. NMSI piloted revisions to the LTF training in Summer 2019 and received overwhelmingly positive feedback from participants. This project explores the effects of the program experience revisions for a cohort of teachers, impacting thousands of middle school students in schools serving military families. This activity has the potential to advance knowledge and understanding around teacher professional development generally, intentional course progression, as well as creating an inclusive college-level coursework accessible school culture.

This study is motivated by two questions that inform the project's overarching goal of improving teacher professional development to increase college-level coursework access and success for underserved students and military connected students:

1. How does the NMSI program influence **student mindsets**, such as STEM course interests, self-identification with STEM subjects, belief in ability to succeed in AP STEM courses, and STEM career interests?
2. How does the NMSI program impact **student achievement** such as measures of critical reasoning for college readiness?

Objectives and Expected Outcomes

Over the course of this project, the implementation and effect of hundreds of participating teachers with access to an enhanced experience, including supports throughout the academic year, on the learning outcomes of thousands of middle school students will be studied.

Figure 2. Objectives and Outcomes by Project Year

Year 1: (AY 2021-22) Cohort 1:	**Objectives:** (1) Train and support at least 100 teachers, serving more than 10,000 students; (2) The training and ongoing support that teachers experience build the skills and mindsets that they need to implement their course with fidelity and develop the mindset that ALL students can be successful in rigorous courses if given the right opportunities and supports; (3) Conduct classroom observations in selected sites and provide ongoing support to participating teachers; and (4) Distribute surveys to teachers and students. **Outcome:** (1) increased student engagement in STEM courses, especially for historically underrepresented students (females, African-American, and Latinx students); (2) increased algebraic thinking ability; (3) improved teacher and student attitudes toward rigorous courses; and (4) increased numbers of underrepresented students who can see themselves in future college-level coursework.
Year 2: (AY 2022-23) Cohort 2	**Objectives:** (1) Train and support at least 100 teachers, serving more than 10,000 students; (2) The training and ongoing support that teachers experience build the skills and mindsets that they need to implement their course with fidelity and develop the mindset that ALL students can be successful in rigorous courses if given the right opportunities and supports; (3) Conduct classroom observations in selected sites and provide ongoing support to participating teachers; and (4) Distribute surveys to teachers and students. **Outcomes**: (1) increased student engagement in STEM courses, especially for historically underrepresented students (females, African-American, and Latinx students); (2) increased algebraic and statistical thinking ability; (3) improved teacher and student attitudes toward rigorous courses; (4) increased numbers of underrepresented students who can see themselves in future college-level coursework.
Year 3: (AY 2023-24) Cohort 3	**Objectives:** (1) Train and support at least 100 teachers, serving more than 10,000 students; (2) The training and ongoing support that teachers experience build the skills and mindsets that they need to implement their course with fidelity and develop the mindset that ALL students can be successful in rigorous courses if given the right opportunities and supports; (3) Conduct classroom observations in selected sites and provide ongoing support to participating teachers; and (4) Distribute surveys to teachers and students. **Outcomes**: (1) increased student engagement in STEM courses, especially for historically underrepresented students (females, African-American, and Latinx); (2) increased algebraic and statistical thinking ability; (3) improved teacher and student attitudes toward rigorous courses; (4) increased numbers of underrepresented students who can see themselves in future

	college-level coursework.

Figure 3 describes the overall logic model for LTF. Reading from left to right, the model displays the inputs, project outcomes, project outputs, and long-term outcomes. The combination of outcomes and objectives will drive long-term shifts in conditions, especially for underserved students, so they are prepared to take rigorous AP and college preparatory coursework.

Figure 3. LTF Logic Model

	Components	Outcomes	Outputs	Fidelity measures	Impacts
Teacher Content Workshops	Pedagogical content and experience-based training that provides clear models of engaging STEM coursework as teachers experience learning from diverse students' perspective.	More teachers receive training, increasing their knowledge and confidence delivering engaging lessons.	Increase in teacher confidence in delivering rigorous, engaging lessons and supporting students.	Teacher Surveys Observation	More teachers are prepared and equipped to support students – especially those who have historically been furthest from engaging learning – to access and succeed in STEM.
3 Year Training Arc	Implementation-driven, collaborative training that leverages research-backed principles to positively impact teacher actions in the classroom via a training arc that reflects teacher learning progressions.	Teachers have an active role in their professional development experience and opportunities to practice instructional strategies and are connected to a networked community of content-focused teachers.	Teachers increasingly build and utilize pedagogical content knowledge and new skills and strategies learned in training in their classrooms.	Teacher Surveys Training Attendance Observation	Teachers implement high-quality, research-backed strategies in their classrooms that are rigorous and student-centered.
Curricular Supports	Online supports, including content-rich lessons with clear outcome goals and suggested teaching strategies and focus on high-leverage conceptual understandings and model providing enrichment and scaffolding to ensure all students are moving towards success.	Teachers leverage vertically aligned materials, providing enrichment and scaffolding to ensure all students are moving towards success.	Teachers provide aligned learning that engages and excites all students, leading to improved STEM mindsets and algebraic thinking.	Student Outcome Data: - Critical Reasoning for College Readiness assessment -Student mindsets	Students are offered learning experiences that are cohesive, coherent and relevant, increasing the depth of knowledge and leading to greater STEM literacy. academic success, college readiness, and college engagement.
Program enablers: - Schools, counselors, teachers, students, and parents value STEM					

- School unit has an explicit focus on equity - STEM coursework offered and delivered equitably - Large percentage of teachers participating in program and implementing changes in practice - Culturally responsive materials support equitable learning experience for all students

Methodology

This study is motivated by two questions that inform the overarching goal of improving LTF teacher professional development to increase student success:

- **RQ1**: How does the training experience in the NMSI program influence **student mindsets**, such as STEM course interests, self-identification with STEM subjects, belief in ability to succeed in AP STEM courses, and STEM career interests?
- **RQ2:** How does the NMSI program impact **student achievement** in the short term as well as the long term? How do gains in achievement vary by student subgroups?

Data Sources

The sources of data for this project include: (a) items from an instrument measuring student interest and motivation in STEM for RQ1, and (b) items from a validated instrument (Brown, Wilson, & Draney, 2021) assessing students' levels of algebraic thinking (PSM) and statistical thinking (DDM) for RQ2.

This approach utilizes a student assessment of algebraic thinking developed in conjunction with the Berkeley Educational Assessment Research Center (BEAR) at the University of California, Berkeley. The BEAR Center, in partnership with NMSI, has been developing psychometrically sound measures of *discipline specific* college readiness aligned with Common Core, AP Content, and NAGB/EPIC research. Both *mathematical problem solving* and *reasoning with data* are skills that show up across many first year applied college courses; we have chosen to develop assessments of these skills as part of an ongoing project aimed at the broad majority of US high school and college students who will take introductory and survey courses in mathematics and data science in their college careers, but who are not destined to have majors in mathematical or mathematics heavy courses (such as data science, statistics and physics).

This work is not focused on the mathematical *mechanics* of algebra or data analysis, but on the application of these constructs in entry-level coursework outside of mathematics and statistics departments. We focus on demanding curricula in which mathematical and statistical thinking are present, introduced, or explicitly taught in a broad range of content areas, from the physical sciences, and engineering, but also social sciences, health sciences, business, and even the humanities. We have, in our initial work (a) laid out construct maps for critical thinking skills, (b) developed initial sets of items to assess these constructs, and (c) are in the process of designing and empirically evaluating an item-bank for use in the field.

Mathematical problem solving has long been a cornerstone of college readiness. Both the SAT and ACT examinations include significant use of problem-solving skills in their mathematics sections, and it is quite common for universities to require students in non-STEM programs to pass a college algebra class of some sort to graduate. Some level of reasoning with data is needed for everyday life. Knowledge and skills that are important just to read the daily news include understanding the nature of the statistics quoted in reports (sample means, margins of error, etc.), reading and interpreting tables or graphs of data, and determining if and how the results of a particular piece of research are generalizable.

In the *Standards for Success* research (Conley, 2005), college professors, especially those of the natural and social sciences, indicated that prerequisite knowledge of basic statistical and mathematical concepts and techniques plays an important role in entry-level courses. Conley reports that students who fail initial coursework because they lack a prerequisite skill will avoid majors in that area of study, "closing off entire avenues of the curriculum and career pathways" (p. 114). It is interesting to note, however, that Conley observed that data analysis skills, as opposed to algebraic skills, are less important in entry-level coursework for mathematics majors, while they are imperative for success in other majors in the natural sciences, social sciences, professional degrees, and even in some parts of the humanities.

Considerable advancements have been made in recent years to identify the skills and knowledge required for students to be considered "college-ready." The BEAR/NMSI collaborative work began with an extensive review of the literature around college readiness in mathematics. This included analyses of some of the most notable collections of college readiness standards such as:

- Common Core State Standards – Mathematics (CCSS-M, National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).
- Knowledge and Skills for University Success (KSUS, Conley, 2005).
- Standards for College Success – Mathematics and Statistics, Adapted for Integrated Curricula (CBSCS, Mathematics and Statistics Standards Advisory Committee, 2007).

Along with these three documents, three other collections of K-12 academic standards for mathematics and statistics were surveyed:

- Mathematics Framework for the National Assessment of Educational Practices (NAEP, National Assessment Governing Board, 2013),
- Principles and Standards for School Mathematics (NCTM, National Council of Teachers of Mathematics, 2000), and
- Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework (Franklin, Kader, Mewborn, Moreno, Peck, Perry, & Scheaffer, 2007).

Finally, we conducted a survey of Advanced Placement® (AP) coursework and extracted elements of algebraic and statistical reasoning that were required as prerequisite in a large variety of coursework (College Board, 2014a, 2014b, 2015, 2016). We used these to begin developing our constructs and our measures. This online measure will be administered to students in the trained teachers' classrooms at the beginning and end of each academic year. This will enable us to estimate the growth in algebraic thinking for students in participating teachers' classrooms and estimate the improvement and progress toward college-readiness algebraic thinking for students in these schools.

This resulted in the development of the Critical Reasoning for College Readiness assessment (CR4CR) which includes three strands: 1) Algebraic Thinking as measured by the Problem Solving using Math (PSM) strand; 2) Statistical Thinking as measured by the Data-based Decision Making (DDM) strand; and 3) Computational Thinking as measured by the CoT strand. This study utilizes measured for only the first two strands, PSM and DDM.

Data

The data for addressing the research questions comes from administering two sets of instruments to students of teachers who participated in the LTF training. The first set addresses RQ1, which focuses on students' mindsets regarding STEM. These mindsets were assessed through four specific questions. These include:

- STEM1: I am confident that I can be successful in the STEM (science, technology, engineering, and math) courses I enrolled in for next year
- STEM2: I am interested in taking more challenging STEM courses
- STEM3: There are many opportunities for me to engage in STEM (either coursework or activities) at my school
- STEM4: I am interested in exploring STEM career options after high school.

These mindset questions were first administered in the Spring of 2022 for Cohort 1, and subsequently in both Fall and Spring for Cohorts 2 and 3. This allowed for the investigation of change in student mindsets over the course of the academic year for these two cohorts. The second set of data addresses RQ2, which focuses on student achievement. This is measured through the CR4CR assessment described above. All students in Cohort 1 (both Math and Science courses) took the PSM assessment in the 2021-22 academic year. Subsequently, NMSI requested that students in middle school Science courses take the DDM measure and that students in middle school math courses take the PSM measure, so the DDM assessment was administered to Science classes in the 2022-23 and 2023-24 academic years.

Student Sample Sizes

For Cohort 1, more than 1700 students were assessed in the Fall with the PSM measure from 29 teachers who had undergone NMSI training. Additionally, over 1400 students were assessed in the Spring from these same teachers. Of those 1134 were tested at both time points. Almost 1600 students provided STEM perceptions in the Spring of 2022 for Cohort 1. This group was comprised of students self-identifying as female (36.5%) male (50.3%), transgender (2.0%), and non-binary (3.6%). Another 4.8% chose not to answer the gender question. This group also self-identified as Black (14.1%), Hispanic (23.6%) and White (42.1%).

Similarly, Cohort 2 provided 1720 students in the Fall from 28 middle school Math courses taking the PSM measure and another 1409 students from 20 middle school Science courses taking the DDM measure. Together these courses provide 3423 student STEM perceptions. This Fall group was made up of students self-identifying as female (41.7%) male (45.8%), transgender (1.1%), and non-binary (2.5%). Another 8.7% chose not to answer the gender question. This group also self-identified as Black (17.8%), Hispanic (32.6%) and White (27.0%). In the Spring, Cohort 2 is comprised of 1354 PSM students and 1127 DDM students, with 2795 students providing STEM perceptions. The Spring group was made up of students self-identifying as female (43.5%) male (46.4%), transgender (1.4%), and non-binary (2.0%). Another 6.6% chose not to answer the gender question. This group also self-identified as Black (16.3%), Hispanic (34.9%) and White (30.7%).

Cohort 3 saw a substantial increase in participation, with 2318 students from 42 LTF-trained teachers taking the PSM in the Fall and 2116 providing PSM data in the Spring. Similarly, 2659 students from 33 teachers provided DDM data in the Fall and 1153 did so in the Spring. For Cohort 3, over 6100 students provided STEM perceptions in the Fall and over 5100 provided STEM perceptions in the Spring. The Fall group was made up of students self-identifying as female (43.7%) male (47.7%), transgender (1.0%), and non-binary (1.3%). Another 6.3% chose not to answer the gender question. This group also self-identified as Black (14.2%), Hispanic (37.9%) and White (27.8%). The Spring group was made up of students self-identifying as female (43.9%) male (47.3%), transgender (1.0%), and non-binary (1.2%). Another 6.5% chose not to answer the gender question. This group also self-identified as Black (14.3%), Hispanic (39.0%) and White (28.9%).

Results

RQ1: Student STEM Perceptions

For RQ1 we found that student confidence in taking STEM courses improved throughout the year for students in classes taught by NMSI-trained teachers, and that this improvement was consistent across student subgroups. We saw some group differences in students' interest in STEM careers by Hispanic students, and for Cohort 2 a greater increase in interest in STEM courses for Black students.

We found that, generally, student STEM perceptions in Cohort 2 were in the mid-range with values of around 2.6-2.7 on average on a four-point (1-4) scale (See Table 3). We found support for the argument that students taught by teachers with NMSI training increased at least their confidence in taking STEM courses. As Table 4 shows, the mean Fall to Spring change was small (.066) but statistically significant ($p < 0.001$). None of the other STEM perceptions changed from Fall to Spring for Cohort 2.

We also explored if gains in STEM perceptions were consistent across student sub-groups, specifically Female, Black, and Hispanic students. We found no significant difference in changes in STEM perceptions for Females students (see Table 6). However, we found a significant differences in changes in STEM perceptions for Black students regarding interest in courses. Black students' interest in taking more STEM courses increased greater (.099) than for non-Black students (-.022) in Cohort 2 (See Table 7). Conversely, we found that Hispanic students in Cohort 2 had a significantly lower increase in interest in STEM careers than non-Hispanic students (-.063 vs .047; $p < 0.04$).

We also explored if gains in STEM perceptions were related to student achievement outcomes our student achievement gains. We found no significant relationships between student achievement in either PSM or DDM at Fall or Spring, or PSM or DDM gains, with improvements in STEM perceptions for Cohort 2 (see Table 9).

We conducted comparable analyses for Cohort 3 and found similar results. STEM perceptions were on average in the mid-range (2.3 to 2.6) on a four-point scale (Table 10). No differences were found for Females or Blacks, and the difference for Hispanics on gain in interest in STEM careers in Cohort 2 was also found in Cohort 3, although less significant ($p < 0.06$; see Table 13).

RQ2: Student Achievement

RQ2 asks about the impact of NMSI-trained teachers on student achievement and the consistency of that achievement across student sub-groups. To address that question, we looked at gains in student achievement across Cohorts and variations of those gains between student subgroups.

We found that student academic achievement did improve throughout the year for students in classes taught by NMSI-trained teachers, and that this improvement was consistent across student subgroups. We saw no group differences in students' increased achievement for Female, Black, and Hispanic students. We also saw greater increases across Cohorts for PSM as compared to DDM, indicating that perhaps the training was more applicable in middle school math course settings than in middle school science course settings.

Cohort 1

We found that students in classes taught by NMSI-trained teachers did demonstrate positive gains in PSM scores for Cohort 1. Paired samples comparisons showed that students increased their PSM scores significantly, with an average of increased of .14 ($t = 17.057$, $df = 1,1133$; $p < 0.001$). This is comparable to a .5 Cohen's d effect size (see Table 15). In addition, when we compared distribution of scores along the waypoints on the PSM learning progression, we found that a greater number of students were performing at the third waypoint (61.9%) in the Spring than in the Fall (51.9%). This difference was also statistically significant ($p < 0.001$). Waypoints reflect qualitatively different levels along the learning progression. We found a strong negative correlation between Fall PSM scores and gains, indicating students scoring lower in Fall tended to gain more over the course of the year ($r = -.376$; $p < 0.001$; see Table 17).

Cohort 2

Cohort 2 provided our first opportunity to look at both gains in PSM and gains in DDM, and we found that students increased in both areas throughout the year but gained more in PSM than in DDM. Paired sample t-tests show that students increased an average of .233 points on the PSM scale and an average of .099 points on the DDM scale (Table 19). This equates to an effect size of .314 for PSM and .122 for DDM (Table 20). We found the Fall and Spring PSM scores were highly correlated ($r = .437$; $p < 0.001$) and that gains were also greatly associated with Fall scores ($r = -.683$; $p < 0.001$; Table 21 and Figure 4) again indicating that students performing lower in the Fall tended to gain more throughout the academic year.

A comparison of PSM waypoints indicates a slight increase in the proportion of students scoring above Waypoint 1 in Spring (64.8%) compared to Fall (55.5%). This increase is statistically significant ($t = 8.26$; $p < 0.001$),Table 22) and indicates an effect size of approximately .25 (Table 23). As mentioned, Cohort 2 also showed increases in DDM from Fall to Spring. The 940 matched students showed an increase of .099 points. These increases were also strongly negatively related to Fall DDM scores, indicating students scoring lower on the DDM assessment in the Fall tended to gain more throughout the year ($r = -.594$; $p < 0.001$). A comparison of DDM waypoints indicates an increase in the proportion of students scoring above Waypoint 2 in Spring (42.0%) compared to Fall (27.3%). This increase is statistically significant ($t = 8.24$; $p < 0.001$; Table 25) and indicates an effect size of approximately .27 (Table 26), slightly higher than the effect size found for PSM with this Cohort.

Cohort 2 Subgroups

As indicated in Tables 27 through 29, there was no significant difference found for changes in either PSM or DDM scores from Fall to Spring in Cohort 2 for Female students, Black students, or Hispanic students. This indicates that the growth of students in classes with NMSI-trained teachers is consistent across gender and ethnic boundaries.

Cohort 3

In Cohort 3 we found that students increased in PSM but dropped in DDM. Paired sample t-tests show that students increased an average of .16 points on the PSM scale and dropped an average of .077 points on the DDM scale (Table 31). This equates to an effect size of .106 for PSM and .129 for DDM (Table 32). As before, we note that growth in PSM is negatively correlated with Fall scores, indicating that students who gained the most were those that started off behind in these classrooms ($r = -.476$; $p < 0.001$). Thus, it would appear the Fall gap between high and low performing students would be reduced.

A comparison of PSM waypoints indicates a slight increase in the proportion of students scoring above Waypoint 1 in Spring (31.5%) compared to Fall (27.1%). This increase is statistically significant ($t = 4.48$; $p < 0.001$; Table 34) and indicates an effect size of approximately .25 (Table 35).

As mentioned, Cohort 3 showed a decrease in DDM from Fall to Spring. The 812 matched students showed an average decrease of .077 points. These increases were also strongly negatively related to Fall DDM scores, indicating students scoring lower on the DDM assessment in the Fall tended to gain more throughout the year ($r = -.605$; $p < 0.001$), consistent with Cohort 2. A comparison of DDM waypoints indicates a slight increase in the proportion of students scoring above Waypoint 2 in Spring (32.3%) compared to Fall (28.8%). This increase is marginally statistically significant ($t = 1.746$; $p < 0.08$; Table 37) and indicates an effect size of only .06 (Table 38).

Cohort 3 Subgroups

As indicated in Tables 39 though 41 below, there was no significant difference found for changes in either PSM or DDM scores from Fall to Spring in Cohort 3 for Female students, Black students, or Hispanic students. This indicates that, as with Cohort 2, the growth of students in classes with LTF-trained teachers is consistent across gender and ethnic boundaries in Cohort 3.

Summary and Conclusion

The project examined the impact of an enhanced professional development program, developed by the National Math and Science Initiative (NMSI), on middle school teachers and their students across three distinct locations. NMSI, a recognized leader in educational reform, refined and expanded its well-established program by introducing additional resources, tools, and year-round support aimed at improving the teaching practices of middle school educators. The core objective was to assess how these enhancements influenced student perceptions of STEM (Science, Technology, Engineering, Mathematics) subjects and their subsequent learning outcomes, particularly in algebraic and statistical thinking.

NMSI’s program, rooted in the belief that exceptional teaching is the key to unlocking student potential, has a longstanding history of driving academic success, especially among underrepresented students. The program provides a three-year progression of training that equips teachers with the necessary skills and content knowledge to raise academic expectations and foster advanced levels of thinking and learning in their students. The program’s design emphasizes inquiry-based learning, instructional scaffolding, and intentional course progression, all of which are aimed at preparing students for college-level coursework and careers in STEM fields.

This study was guided by two primary research questions: (1) How does the NMSI training influence student mindsets, such as their interest in STEM courses, self-identification with STEM subjects, belief in their ability to succeed in Advanced Placement (AP) STEM courses, and interest in STEM careers? (2) How does the NMSI training impact student achievement, particularly in algebraic and statistical thinking, and how do these gains vary across different student subgroups?

To answer these questions, data was collected from students taught by NMSI-trained teachers across three cohorts over three academic years (2021-2024). The assessments focused on student perceptions of STEM and their achievement in key areas of algebraic and statistical thinking. The study revealed that students in classrooms led by NMSI-trained teachers showed significant improvements in their confidence in taking STEM courses, with measurable gains in both algebraic and statistical thinking, particularly in algebraic reasoning. These gains were consistent across all three cohorts and across diverse student groups, including females, African-American, and Hispanic students.

In addition to improved student confidence in STEM, the study found that interest in STEM careers increased significantly for Black students in Cohort 2, although Hispanic students showed less increase in interest compared to non-Hispanic students. Importantly, the study also revealed that the gains in student achievement were consistent across gender and ethnic boundaries, indicating that the program effectively supports diverse student populations.

The results of this project underscore the effectiveness of NMSI's enhanced program in positively influencing both student mindsets and academic achievement in STEM subjects. By providing middle school teachers with a comprehensive, practice-based professional development experience, the program equipped educators with the tools and strategies necessary to improve student outcomes significantly. The program's emphasis on content-rich training, instructional scaffolding, and intentional course progression contributed to notable improvements in students' confidence in their ability to succeed in STEM courses, as well as their actual performance in algebraic and statistical thinking.

The study's findings are particularly encouraging given the consistent gains observed across diverse student populations, including those traditionally underrepresented in rigorous academic courses. This indicates that the program not only helps to close achievement gaps but also fosters an inclusive learning environment where all students have the opportunity to excel in STEM fields. The increases in interest in STEM careers among Black students and the consistent academic gains across all cohorts further highlight the potential of the program to inspire and prepare the next generation of STEM professionals.

Overall, this project demonstrates the significant impact that well-structured, sustained professional development can have on both teachers and students. By investing in the ongoing development of educators, NMSI's program helps to create a pipeline of students who are not only prepared for college-level coursework but are also motivated to pursue STEM careers. As educational institutions and policymakers continue to seek ways to improve STEM education and close achievement gaps, the findings from this project offer valuable insights into the role of professional development in achieving these goals. The success of the program serves as a model for other initiatives aimed at enhancing teacher effectiveness and improving student outcomes in critical academic areas.

Recommended Additional Research and Next Steps

Based on the findings of this evaluation, several areas of additional research can be explored to further refine and enhance the program's impact on teacher development and student outcomes in STEM education. These next steps aim to address current limitations, explore new areas of potential impact, and optimize the program for diverse educational environments.

Longitudinal Studies on Student Success

While this study captured the impact of the NMSI program over three academic years, a longer-term longitudinal study could examine the sustained effects of the program on students as they progress through higher education and into STEM careers. This research could explore the impact of NMSI-trained teachers on students' college readiness and their success in Advanced Placement (AP) or college-level STEM courses. In addition, career trajectories of students exposed to NMSI-trained teaching, specifically examining how early interventions influence the pursuit of STEM-related degrees and professions could be examined. Further, tracking postsecondary achievement (e.g., college graduation rates, STEM degree completion) to determine if the program leads to long-term academic success and career readiness is a worthwhile pursuit.

Deeper Exploration of Subgroup Variations

The report indicated differing impacts on various student subgroups, particularly Black and Hispanic students. Future research could investigate the factors driving differential impacts in STEM career aspirations between Hispanic students and other subgroups. Understanding cultural, social, or educational influences that may shape these outcomes could help tailor the program for greater inclusivity. Further study could also explore gender-specific outcomes in greater depth, assessing whether targeted interventions might be needed to further close the gender gap in STEM career interest and achievement.

Teacher Retention and Professional Growth

Another area worth exploring is the long-term impact of NMSI training on teachers themselves, particularly in terms of teacher retention rates, such as whether participation in sustained professional development programs like LTF affects long-term teacher retention, particularly in STEM subjects where teacher turnover is often high. Also, career progression could be explored, such as how NMSI training influences teachers' professional growth, leadership opportunities, and further educational development. Lastly, exploring how teacher attitudes, efficacy, and engagement change after participating, and whether their practices continue to evolve beyond the program's three-year structure would be interesting.

Comparative Studies with Other Professional Development Models

Conducting comparative studies between NMSI and other prominent STEM-focused professional development programs could provide insights into best practices across programs and identifying unique features of the NMSI program that yield superior outcomes. It would also provide an opportunity to compare the cost-effectiveness and scalability versus other professional development initiatives. Exploring the efficacy of hybrid or virtual training models for delivering the NMSI program, especially in light of the increasing importance of remote learning and professional development in education is another interesting topic.

Customization for Local Contexts and Diverse Learning Environments

Future research could focus on how to best adapt the program for diverse educational settings, particularly understanding how urban vs. rural settings or low-resource schools impact the implementation and outcomes. Identifying specific barriers or enablers in different contexts could inform program adjustments. Investigating customization of NMSI training for specific regions or cultural contexts could improve its effectiveness for underrepresented groups. In addition, piloting context-specific modules that address unique challenges faced by educators in high-need schools, such as strategies for managing larger class sizes or overcoming resource constraints may prove beneficial.

STEM Engagement Beyond the Classroom

This report highlights increased STEM confidence and interest among students, but more research is needed to explore how these changes translate into STEM engagement outside of school. Future research could explore: the effectiveness of extracurricular STEM programs (e.g., science fairs, coding clubs, internships) in reinforcing the skills and mindsets developed through the program; ways in which community partnerships with STEM industries, universities, or local organizations could provide practical experiences and mentorship opportunities to complement the classroom-based program; and whether integrating project-based learning and real-world applications into the NMSI curriculum can increase students' engagement and interest in STEM fields.

Impact of Culturally Responsive Teaching on STEM Outcomes

Given the emphasis on culturally responsive teaching in the NMSI program, further research could delve into how this aspect influences student success, particularly among marginalized groups. This could include assessing the long-term effects of culturally responsive pedagogy on student achievement in STEM, particularly for historically underserved groups and examining how culturally relevant curriculum materials and teaching strategies can be further refined to enhance student engagement and performance in STEM subjects.

References

Ball, D. L. & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes and L. Darling-Hammond (Eds.), Teaching as the learning profession: Handbook of policy and practice (pp. 3-32). San Francisco: Jossey Bass.

Birman, B. S., Le Floch, K. C., Kelkotka, A., Ludwig, M., Taylor, J., Walters, K., . . . Yoon, K. (2007). State and local implementation of the No Child Left Behind Act: Vol. 2. Teacher quality under NCLB: Interim report. Washington, DC: U.S. Department of Education; Office of Planning, Evaluation and Policy Development; Policy and Program Studies Service.

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick.* Harvard University Press.

Brown, R., & Phelan, J. (2017). *ALSDE/A+ College Ready LTF Teacher Implementation Evaluation Study*. Retrieved from http://www.westcoastanalytics.com/uploads/6/9/6/7/69675515/wca_ltf_report_final_2017.pdf.

Brown, R., Wilson, M., & Draney, K. (2021). Measuring Statistical and Algebraic Thinking Using Simulations. In Using Cognitive and Affective Metrics in Education-Based Simulations and Games (Vol 2). H.F. O'Neil, E.L. Baker, R.S. Perez, & S. E. Watson, (Eds). Routledge/Taylor & Francis.

Mathematics and Statistics Standards Advisory Committee. (2007). *College Board Standards for College SuccessTM: Mathematics and Statistics, Adapted for Integrated Curricula*. New York, NY: College Board.

College Board. (2013). *AP score distributions—All subjects 1993-2013.* Retrieved from http://research.collegeboard.org/programs/ap/data/archived/2013

College Board. (2014a). *Government and politics comparative: Course description*. New York: The College Board. Available at http://apcentral.collegeboard.com/apc/public/courses/descriptions/index.html

College Board. (2014b). *Government and politics United States: Course description.* New York: The College Board. Available at http://apcentral.collegeboard.com/apc/public/courses/descriptions/index.html

College Board. (2015). *AP Biology: Course and exam description.* New York: The College Board. Available at http://apcentral.collegeboard.com/apc/public/courses/descriptions/index.html

College Board. (2016). *AP Computer Science Principles: Course and Exam Description, including the curriculum framework*. New York: The College Board. Available at http://apcentral.collegeboard.com/apc/public/courses/descriptions/index.html

Conley, D. T. (2005). *College Knowledge: What It Really Takes for Students to Succeed and What We Can Do to Get Them Ready. San Francisco, CA: Jossey-Bass*

College Board. (2013). *AP score distributions—All subjects 1993-2013.* Retrieved from http://research.collegeboard.org/programs/ap/data/archived/2013

Corcoran, T., McVay, S., & Riordan, K. (2003). Getting it right: The MISE approach to professional development. Philadelphia: Consortium for Policy Research in Education.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards – Mathematics. Washington, DC.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development.

Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: what matters? *How Teachers Learn, 66*(5), 46-53.

Dehejia, J. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American Statistical Association, 98, 1053-1062.

Ericsson, A. & Pool, R. (2016) Peak: Secrets from the New Science of Expertise. Houghton Mifflin Harcourt: Boston.

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*. *103*(6), 1013-55.

Ferrini-Mundy, Burrill, & Schmidt. (2007). Building teacher capacity for implementing curricular coherence: Mathematics teacher professional development tasks. *Journal of Mathematics Teacher Education*, *10*(4), 311-324.

Fischer, C., Fishman, B., Dede, C., Eisenkraft, A., Frumin, K., Foster, B., Lawrenz, F., Levy, A. J., & McCoy, A. (2018). Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform. *Teaching and Teacher Education, 72*, 107-121.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K- 12 Curriculum Framework. Alexandria, VA: American Statistical Association.

Friedrichsen, P., & Berry, A. (2015). Science teacher PCK learning progressions. In Berry, A., Friedrichsen, P., & Loughran, J. (Eds.), *Re-examining pedagogical content knowledge in science education*. New York, NY: Routledge.

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.

Gay, G. (2010). Culturally responsive teaching: Theory, research, and practice (2nd ed.). New York: Teachers College Press.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. Teachers and Teaching: theory and practice, 15(2), 273-289.

Guskey, T. R. (2016). Gauge impact with 5 levels of data. SMEC2016 Organising Committee, 6.

Hammond, Z. (2015). Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students. Thousand Oaks, CA: Corwin, a SAGE Company.

Hiebert, J. (1999). Relationships betweeen research and the NCTM standards. Journal for Research in Mathematics Education, v30 n1, 3-19.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B. 53, 597-610.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. Evaluation Review, 23, 418-444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. Multivariate Behavioral Research, 36, 249-277.

MacKinnon, D. P. (2008). Introduction to statistical mediation analysis. New York, NY: Lawrence Erlbaum Associates.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychological Methods, 9, 403-425.

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), Longitudinal data analysis (pp. 143-165). Boca Raton, FL: Chapman & Hall/CCM.

National Assessment of Educational Practices Governing Board. (2013). *Mathematics Framework for the 2013 National Assessment of Educational Progress*. Washington D.C.: U.S. Department of Education.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: The National Council of Teachers of Mathematics, Inc.

Pella, S. (2015). Pedagogical reasoning and action: Affordances of practice-based teacher professional development. *Teacher Education Quarterly,* 81-101.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. Structural Equation Modeling, 18, 161-182.

Rosenbaum, P. R. (2002). *Observational Studies*, 2nd ed. Springer, New York.

Rosenbaum, P.R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*, 33-38.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores, relating theory to practice. Biometrics, 52, 249-264.

Rubin, D.B. 2007. "The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials." *Statistical Medicine (26)*, 20-36.

Schneider, R. M., & Plasman, K. (2011). Science Teacher Learning Progressions. Review of Educational Research, 81(4), 530–565.

Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. Statistical Science, 25(1), 1-21.

Supovitz, J.A., & Turner, H.M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(9), 963–980.

U.S. Department of Education. (2013). National Center for Education Statistics, Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey," SY 2012–13 Provisional Version 1a. Available at http://nces.ed.gov/ccd/ccddata.asp

Yoon, K. S., Duncan, T., Wen-Yu, S., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007 – No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Winitzky, N., & Arends, R. (1991). Translating research into practice: The effects of various forms of training and clinical experience on preservice students; knowledge, skill, and reflectiveness. Journal of Teacher Education, 42(1), 52-65.

Appendix

Table 1. Student Sample Sizes

Cohort	Fall PSM	Spring PSM	Fall DDM	Spring DDM	Fall STEM Perceptions	Spring STEM Perceptions
1 (21-22)	1720	1479	N/A	N/A	N/A	1595
2 (22-23)	1723	1354	1409	1127	3423	2795
3 (23-24)	2318	2116	2659	1153	6193	5129

Table 2. Fall and Spring Matched Files Student Sample Sizes by Cohort

Cohort	PSM	DDM
1 (21-22)	1134	N/A
2 (22-23)	1123	940
3 (23-24)	1516	812

Table 3. Cohort 2 Student STEM Perceptions

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Confident	3235	1.00	4.00	2.6597	.01650	.93847
Interested Courses	3234	1.00	4.00	2.3800	.01793	1.01944
Opportunities	3234	1.00	4.00	2.6308	.01722	.97940
Interested Career	3233	1.00	4.00	2.3724	.01904	1.08255
Confident	2712	1.00	4.00	2.7124	.01798	.93651
Interested Courses	2712	1.00	4.00	2.3569	.01954	1.01782
Opportunities	2711	1.00	4.00	2.6278	.01858	.96764
Interested Career	2711	1.00	4.00	2.3412	.02047	1.06563
Valid N (listwise)	2247					

Table 4. Cohort 2 Gains in STEM Perceptions

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
STEM1Gain	2249	-3.00	3.00	.0587	.02079	.98617
STEM2Gain	2249	-3.00	3.00	-.0116	.02143	1.01626
STEM3Gain	2248	-3.00	3.00	.0125	.02324	1.10177
STEM4Gain	2247	-3.00	3.00	-.0040	.02413	1.14391
Valid N (listwise)	2247					

Table 5. Cohort 2 Paired Sample T-test STEM Perceptions

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	Confident - Confident	.06571	.99041	.02161	.02333	.10810	3.041	2099	.001	.002
Pair 2	Interested Courses - Interested Courses	-.00333	1.01183	.02208	-.04663	.03997	-.151	2099	.440	.880
Pair 3	Opportunities - Opportunities	.01239	1.10154	.02404	-.03476	.05954	.515	2098	.303	.606
Pair 4	Interested Career - Interested Career	.01001	1.14782	.02506	-.03913	.05915	.399	2097	.345	.690

Table 6. Cohort 2 Change in STEM Perceptions by Gender

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
STEM1Gain * Female	Between Groups	(Combined)	.981	1	.981	1.000	.317
	Within Groups		2057.950	2098	.981		
	Total		2058.931	2099			
STEM2Gain * Female	Between Groups	(Combined)	.159	1	.159	.155	.694
	Within Groups		2148.818	2098	1.024		
	Total		2148.977	2099			
STEM3Gain * Female	Between Groups	(Combined)	2.593	1	2.593	2.138	.144
	Within Groups		2543.085	2097	1.213		
	Total		2545.678	2098			
STEM4Gain * Female	Between Groups	(Combined)	.009	1	.009	.007	.933
	Within Groups		2762.780	2096	1.318		
	Total		2762.790	2097			

Table 7. Cohort 2 Change in STEM Perceptions for Black Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
STEM1Gain * Black	Between Groups	(Combined)	.330	1	.330	.336	.562
	Within Groups		2058.602	2098	.981		
	Total		2058.931	2099			
STEM2Gain * Black	Between Groups	(Combined)	3.957	1	3.957	3.870	.049
	Within Groups		2145.020	2098	1.022		
	Total		2148.977	2099			
STEM3Gain * Black	Between Groups	(Combined)	3.052	1	3.052	2.517	.113
	Within Groups		2542.626	2097	1.213		
	Total		2545.678	2098			
STEM4Gain * Black	Between Groups	(Combined)	.143	1	.143	.108	.742
	Within Groups		2762.647	2096	1.318		
	Total		2762.790	2097			

Table 8. Cohort 2 Change in STEM Perceptions for Hispanic Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
STEM1Gain * HispanicBinary	Between Groups	(Combined)	.271	1	.271	.276	.599
	Within Groups		2058.660	2098	.981		
	Total		2058.931	2099			
STEM2Gain * HispanicBinary	Between Groups	(Combined)	.029	1	.029	.028	.867
	Within Groups		2148.948	2098	1.024		
	Total		2148.977	2099			
STEM3Gain * HispanicBinary	Between Groups	(Combined)	.185	1	.185	.152	.697
	Within Groups		2545.493	2097	1.214		
	Total		2545.678	2098			
STEM4Gain * HispanicBinary	Between Groups	(Combined)	5.572	1	5.572	4.236	.040
	Within Groups		2757.217	2096	1.315		
	Total		2762.790	2097			

Table 9. Cohort 2 Correlations Among STEM Gains and Student Achievement

Correlations

		PSMFall22	PSMSpring23	DDMFall22	DDMSpring23	PSMGain	DDMGain
STEM1Gain	Pearson Correlation	-.001	.035	-.044	-.016	.030	.024
	Sig. (2-tailed)	.963	.245	.175	.633	.320	.459
	N	1111	1111	933	933	1111	933
STEM2Gain	Pearson Correlation	.001	.020	-.012	-.036	.015	-.020
	Sig. (2-tailed)	.975	.513	.711	.268	.618	.535
	N	1111	1111	933	933	1111	933
STEM3Gain	Pearson Correlation	-.001	.058	.002	-.031	.048	-.028
	Sig. (2-tailed)	.980	.053	.952	.337	.110	.390
	N	1111	1111	933	933	1111	933
STEM4Gain	Pearson Correlation	.032	.030	.039	-.016	-.009	-.046
	Sig. (2-tailed)	.290	.316	.237	.620	.776	.157
	N	1111	1111	933	933	1111	933

Table 10. Cohort 3 STEM Perceptions

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
STEM1Fall23R	5980	1.00	4.00	2.6684	.01184	.91533
STEM2Fall23R	5979	1.00	4.00	2.3424	.01277	.98739
STEM3Fall23R	5978	1.00	4.00	2.6561	.01202	.92904
STEM4Fall23R	5980	1.00	4.00	2.3612	.01372	1.06061
STEM1Spring24R	4909	1.00	4.00	2.7598	.01337	.93643
STEM2Spring24R	4909	1.00	4.00	2.3532	.01429	1.00117
STEM3Spring24R	4908	1.00	4.00	2.6532	.01347	.94367
STEM4Spring24R	4908	1.00	4.00	2.3026	.01500	1.05055
Valid N (listwise)	3894					

Table 11. Cohort 3 Change in STEM Perceptions by Gender

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
STEM1Gain * Female	Between Groups	(Combined)	2.003	1	2.003	2.238	.135
	Within Groups		2970.309	3319	.895		
	Total		2972.311	3320			
STEM2Gain * Female	Between Groups	(Combined)	.107	1	.107	.101	.750
	Within Groups		3509.232	3319	1.057		
	Total		3509.339	3320			
STEM3Gain * Female	Between Groups	(Combined)	.000	1	.000	.000	.996
	Within Groups		3798.542	3318	1.145		
	Total		3798.542	3319			
STEM4Gain * Female	Between Groups	(Combined)	.806	1	.806	.597	.440
	Within Groups		4477.736	3318	1.350		
	Total		4478.542	3319			

Table 12. Cohort 3 Change in STEM Perceptions for Black Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
STEM1Gain * Black	Between Groups	(Combined)	.662	1	.662	.738	.390
	Within Groups		3171.434	3532	.898		
	Total		3172.096	3533			
STEM2Gain * Black	Between Groups	(Combined)	.280	1	.280	.264	.607
	Within Groups		3751.212	3532	1.062		
	Total		3751.492	3533			
STEM3Gain * Black	Between Groups	(Combined)	.065	1	.065	.057	.812
	Within Groups		4044.362	3531	1.145		
	Total		4044.427	3532			
STEM4Gain * Black	Between Groups	(Combined)	.000	1	.000	.000	.994
	Within Groups		4762.278	3531	1.349		
	Total		4762.278	3532			

Table 13. Cohort 3 Change in STEM Perceptions for Hispanic Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
STEM1Gain * HispanicBinary	Between Groups	(Combined)	5.194	1	5.194	5.793	.016
	Within Groups		3166.903	3532	.897		
	Total		3172.096	3533			
STEM2Gain * HispanicBinary	Between Groups	(Combined)	1.370	1	1.370	1.290	.256
	Within Groups		3750.122	3532	1.062		
	Total		3751.492	3533			
STEM3Gain * HispanicBinary	Between Groups	(Combined)	1.933	1	1.933	1.689	.194
	Within Groups		4042.493	3531	1.145		
	Total		4044.427	3532			
STEM4Gain * HispanicBinary	Between Groups	(Combined)	4.811	1	4.811	3.571	.059
	Within Groups		4757.467	3531	1.347		
	Total		4762.278	3532			

Table 14. Cohort 1 Pre-Post PSM Scores

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	PostEAP - PreEAP	1.43705E-001	2.83715E-001	8.42511E-003	1.27175E-001	1.60236E-001	17.057	1133	<.001	<.001

Table 15. Cohort 1 Achievement Increase Effect Size for PSM

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	PostEAP - PreEAP	Cohen's d	.28371	.507	.445	.568
		Hedges' correction	.28390	.506	.444	.568

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 16. Cohort 1 Comparison on Waypoints Fall to Spring

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	SpringWaypoint - FallWaypoint	.10406	.61010	.01812	.06851	.13960	5.744	1133	<.001	<.001

Table 17. Cohort 1 Correlations Between PSM Scores and Growth

Correlations

		PreEAP	PostEAP	GAIN
PreEAP	Pearson Correlation	1	.465**	-.376**
	Sig. (2-tailed)		<.001	<.001
	N	1134	1134	1134
PostEAP	Pearson Correlation	.465**	1	.645**
	Sig. (2-tailed)	<.001		<.001
	N	1134	1134	1134
GAIN	Pearson Correlation	-.376**	.645**	1
	Sig. (2-tailed)	<.001	<.001	
	N	1134	1134	1134

**. Correlation is significant at the 0.01 level (2-tailed).

Table 18. Cohort 2 PSM and DDM Paired Sample Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PSMSpring23	-.8108	1123	.60328	.01800
	PSMFall22	-1.0442	1123	.77048	.02299
Pair 2	DDMSpring23	-1.1970	940	.68408	.02231
	DDMFall22	-1.2959	940	.68438	.02232

Table 19. Cohort 2 Paired Samples Test

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	PSMSpring23 - PSMFall22	.23340	.74236	.02215	.18994	.27687	10.536	1122	<.001	<.001
Pair 2	DDMSpring23 - DDMFall22	.09885	.81220	.02649	.04686	.15084	3.731	939	<.001	<.001

Table 20. Cohort 2 Achievement Increase Effect Size for PSM and DDM

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
Pair 1	PSMSpring23 - PSMFall22	Cohen's d	.74236	.314	.254	.374
		Hedges' correction	.74286	.314	.254	.374
Pair 2	DDMSpring23 - DDMFall22	Cohen's d	.81220	.122	.058	.186
		Hedges' correction	.81285	.122	.057	.186

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 21. Cohort 2 Correlations Among Fall, Spring, and Gain Scores for PSM

Correlations

		WLEFall22	WLESpring23	Gain
WLEFall22	Pearson Correlation	1	.437**	-.683**
	Sig. (2-tailed)		<.001	<.001
	N	1123	1123	1123
WLESpring23	Pearson Correlation	.437**	1	.359**
	Sig. (2-tailed)	<.001		<.001
	N	1123	1123	1123
Gain	Pearson Correlation	-.683**	.359**	1
	Sig. (2-tailed)	<.001	<.001	
	N	1123	1123	1123

**. Correlation is significant at the 0.01 level (2-tailed).

Table 22. Cohort 2 Comparison on PSM Waypoints Fall to Spring

Paired Samples Test

		Paired Differences							Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	SpringPSMWaypoint - FallPSMWaypoint	.20926	.84852	.02532	.15958	.25894	8.264	1122	<.001	<.001

Table 23. Cohort 2 PSM Waypoint Increase Effect Size

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	SpringPSMWaypoint - FallPSMWaypoint	Cohen's d	.84852	.247	.187	.306
		Hedges' correction	.84909	.246	.187	.306

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 24. Cohort 2 Correlations Among Fall, Spring, and Gain Scores for DDM

Correlations

		WLEFall22	WLESpring23	Gain
WLEFall22	Pearson Correlation	1	.295**	-.594**
	Sig. (2-tailed)		<.001	<.001
	N	940	940	940
WLESpring23	Pearson Correlation	.295**	1	.593**
	Sig. (2-tailed)	<.001		<.001
	N	940	940	940
Gain	Pearson Correlation	-.594**	.593**	1
	Sig. (2-tailed)	<.001	<.001	
	N	940	940	940

**. Correlation is significant at the 0.01 level (2-tailed).

Table 25. Cohort 2 Comparison on DDM Waypoints Fall to Spring

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	SpringDDMWaypoint - FallDDMWaypoint	.18085	.67296	.02195	.13777	.22393	8.239	939	<.001	<.001

Table 26. Cohort 2 Waypoint Increase Effect Size for DDM

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	SpringDDMWaypoint - FallDDMWaypoint	Cohen's d	.67296	.269	.204	.334
		Hedges' correction	.67350	.269	.203	.333

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 27. Cohort 2 Change in Student Achievement for Female Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PSMGain * Female	Between Groups	(Combined)	.274	1	.274	.496	.481
	Within Groups		618.057	1121	.551		
	Total		618.330	1122			
DDMGain * Female	Between Groups	(Combined)	.124	1	.124	.188	.665
	Within Groups		619.312	938	.660		
	Total		619.436	939			

Table 28. Cohort 2 Change in Student Achievement for Black Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PSMGain * Black	Between Groups	(Combined)	.005	1	.005	.010	.922
	Within Groups		618.325	1121	.552		
	Total		618.330	1122			
DDMGain * Black	Between Groups	(Combined)	1.978	1	1.978	3.005	.083
	Within Groups		617.458	938	.658		
	Total		619.436	939			

Table 29. Cohort 2 Change in Student Achievement for Hispanic Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PSMGain * HispanicBinary	Between Groups	(Combined)	.296	1	.296	.537	.464
	Within Groups		618.034	1121	.551		
	Total		618.330	1122			
DDMGain * HispanicBinary	Between Groups	(Combined)	.500	1	.500	.757	.384
	Within Groups		618.936	938	.660		
	Total		619.436	939			

Table 30. Cohort 3 PSM and DDM Paired Sample Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PSMWLES24	-1.3329238973	1516	1.4570869085	.03742276355
	PSMWLEF23	-1.4912248700	1516	1.3426439135	.03448349265
Pair 2	DDMWLES24	-1.1822627777	812	.49492261477	.01736837887
	DDMWLEF23	-1.1054707686	812	.49831503396	.01748742944

Table 31. Cohort 3 Paired Samples Test

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	PSMWLES24 - PSMWLEF23	.15830097274	1.4937363081	.03836404015	.08304871599	.23355322948	4.126	1515	<.001	<.001
Pair 2	DDMWLES24 - DDMWLEF23	-.07679200908	.59759900148	.02097161366	-.11795705110	-.03562696705	-3.662	811	<.001	<.001

Table 32. Cohort 3 Achievement Increase Effect Size for PSM and DDM

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	PSMWLES24 - PSMWLEF23	Cohen's d	1.4937363081	.106	.055	.156
		Hedges' correction	1.4944762903	.106	.055	.156
Pair 2	DDMWLES24 - DDMWLEF23	Cohen's d	.59759900148	-.129	-.198	-.059
		Hedges' correction	.59815236236	-.128	-.197	-.059

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 33. Cohort 3 Correlations Among Fall, Spring, and Gain Scores for PSM

Correlations

		WLE	WLESpring	WLEGain
WLE	Pearson Correlation	1	.433**	-.476**
	Sig. (2-tailed)		<.001	<.001
	N	2318	1516	1516
WLESpring	Pearson Correlation	.433**	1	.586**
	Sig. (2-tailed)	<.001		<.001
	N	1516	2116	1516
WLEGain	Pearson Correlation	-.476**	.586**	1
	Sig. (2-tailed)	<.001	<.001	
	N	1516	1516	1516

**. Correlation is significant at the 0.01 level (2-tailed).

Table 34. Cohort 3 Comparison on PSM Waypoints Fall to Spring

Paired Samples Test

		Paired Differences Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	Significance One-Sided p	Two-Sided p
Pair 1	SpringPSMWaypoint - FallPSMWaypoint	.15172	1.31884	.03387	.08527	.21816	4.479	1515	<.001	<.001

Table 35. Cohort 3 PSM Waypoint Increase Effect Size

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	SpringPSMWaypoint - FallPSMWaypoint	Cohen's d	1.31884	.115	.065	.166
		Hedges' correction	1.31950	.115	.064	.165

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 36. Cohort 3 Correlations Among Fall, Spring, and Gain Scores for DDM

Correlations

		WLE	WLESpring	GainWLE
WLE	Pearson Correlation	1	.276**	-.605**
	Sig. (2-tailed)		<.001	<.001
	N	2659	812	812
WLESpring	Pearson Correlation	.276**	1	.598**
	Sig. (2-tailed)	<.001		<.001
	N	812	1153	812
GainWLE	Pearson Correlation	-.605**	.598**	1
	Sig. (2-tailed)	<.001	<.001	
	N	812	812	812

**. Correlation is significant at the 0.01 level (2-tailed).

Table 37. Cohort 3 Comparison on DDM Waypoints Fall to Spring

Paired Samples Test

		Paired Differences Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	Significance One-Sided p	Two-Sided p
Pair 1	SpringDDMWaypoint - FallDDMWaypoint	.03695	.60300	.02116	-.00459	.07848	1.746	811	.041	.081

Table 38. Cohort 3 DDM Waypoint Increase Effect Size

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	SpringDDMWaypoint - FallDDMWaypoint	Cohen's d	.60300	.061	-.008	.130
		Hedges' correction	.60356	.061	-.008	.130

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Table 39. Cohort 3 Change in Student Achievement for Female Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PSMWLEGain2324 * Female	Between Groups	(Combined)	.084	1	.084	.038	.846
	Within Groups		3180.200	1431	2.222		
	Total		3180.284	1432			
DDMGainWLE2324 * Female	Between Groups	(Combined)	2.044	1	2.044	5.804	.016
	Within Groups		264.547	751	.352		
	Total		266.591	752			

Table 40. Cohort 3 Change in Student Achievement for Black Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PSMWLEGain2324 * Black	Between Groups	(Combined)	1.721	1	1.721	.771	.380
	Within Groups		3378.620	1514	2.232		
	Total		3380.341	1515			
DDMGainWLE2324 * Black	Between Groups	(Combined)	.012	1	.012	.035	.853
	Within Groups		289.616	810	.358		
	Total		289.628	811			

Table 41. Cohort 3 Change in Student Achievement for Hispanic Students

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PSMWLEGain2324 * HispanicBinary	Between Groups	(Combined)	.135	1	.135	.060	.806
	Within Groups		3380.206	1514	2.233		
	Total		3380.341	1515			
DDMGainWLE2324 * HispanicBinary	Between Groups	(Combined)	.111	1	.111	.310	.578
	Within Groups		289.517	810	.357		
	Total		289.628	811			

PSM Cutpoints

Cutpoints

Fall 2022 PSM Proficiency Reports

Level Name	Lower Bound	Upper Bound
PSM4	1.0	3.0
PSM3	0.0	1.0
PSM2	-1.1	0.0
PSM1	-2.0	-1.1
PSM0	-5.0	-2.0

DDM Cutpoints

Cutpoints

Level Name	Lower Bound	Upper Bound
DDM5	1.805	5.0
DDM4	0.395	1.805
DDM3	-0.935	0.395
DDM2	-2.425	-0.935
DDM1	-5.0	-2.425

Figure 4. Cohort 2 Correlations between Fall and Spring PSM Scores

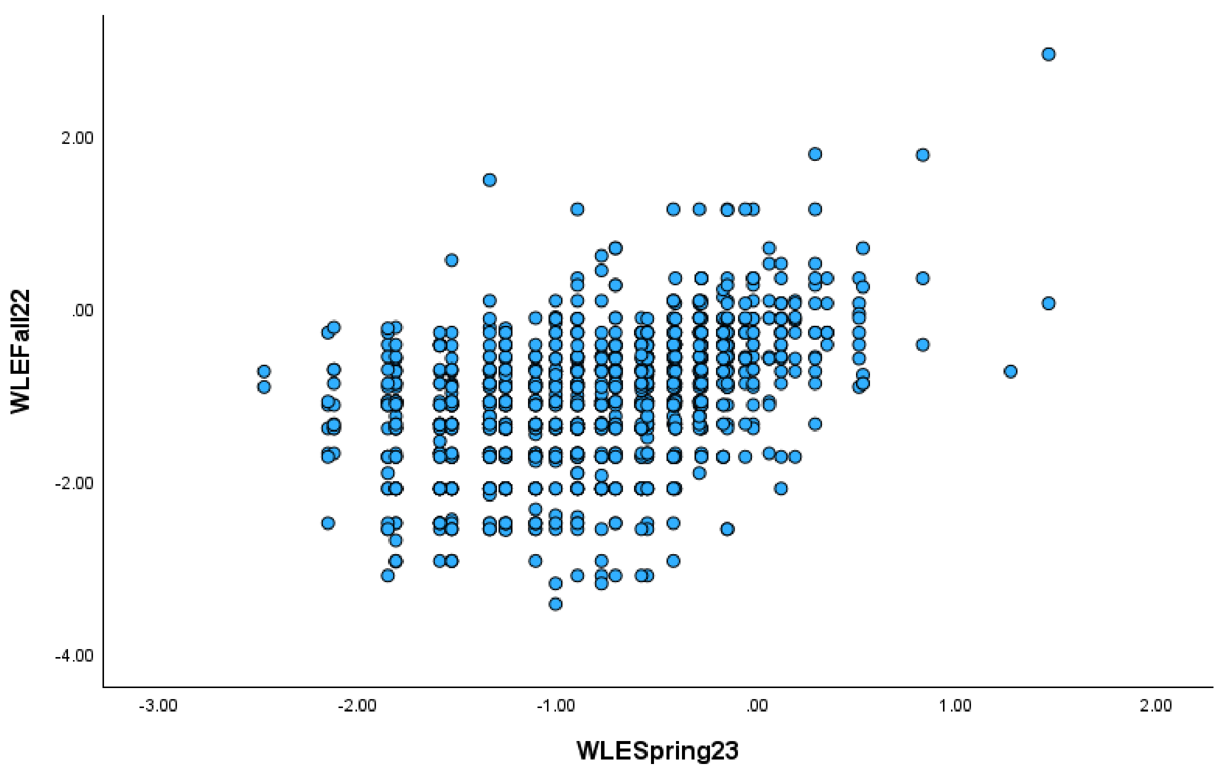

Figure 5. Cohort 2 Correlations between Fall and Spring DDM Scores

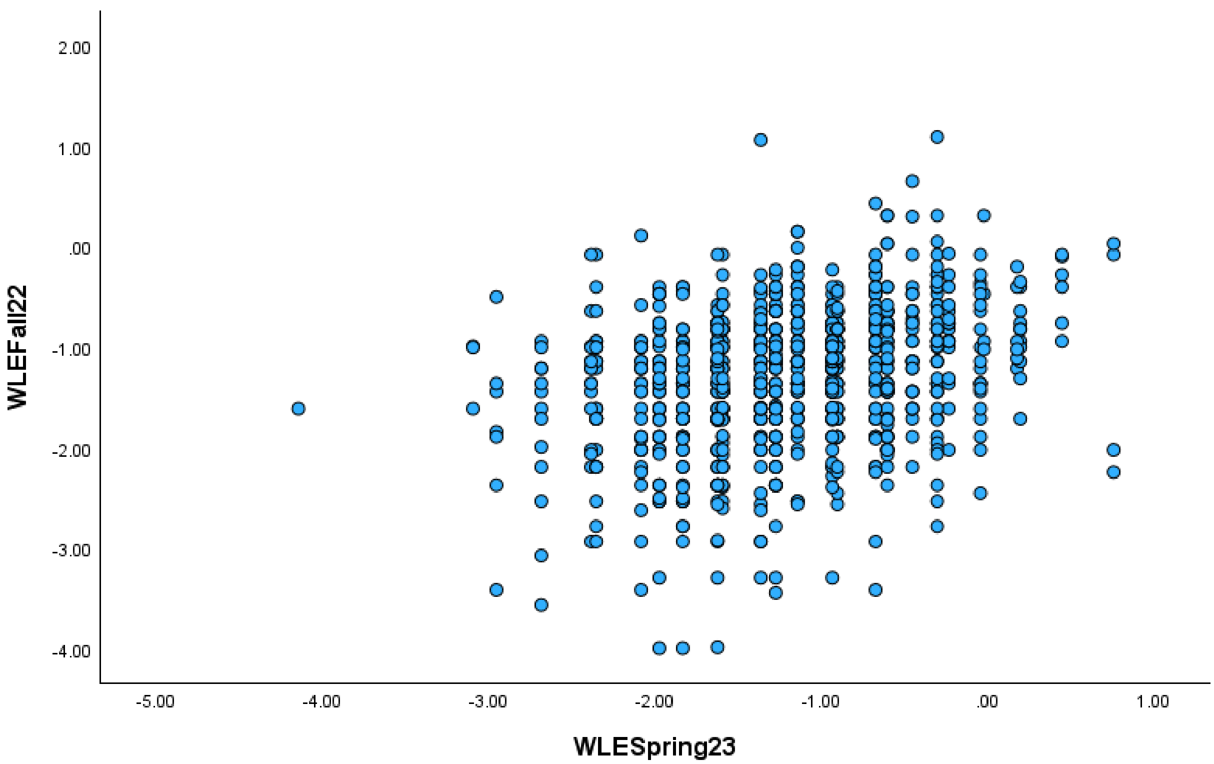

Figure 6. Cohort 3 Correlations between Fall and Spring PSM Scores

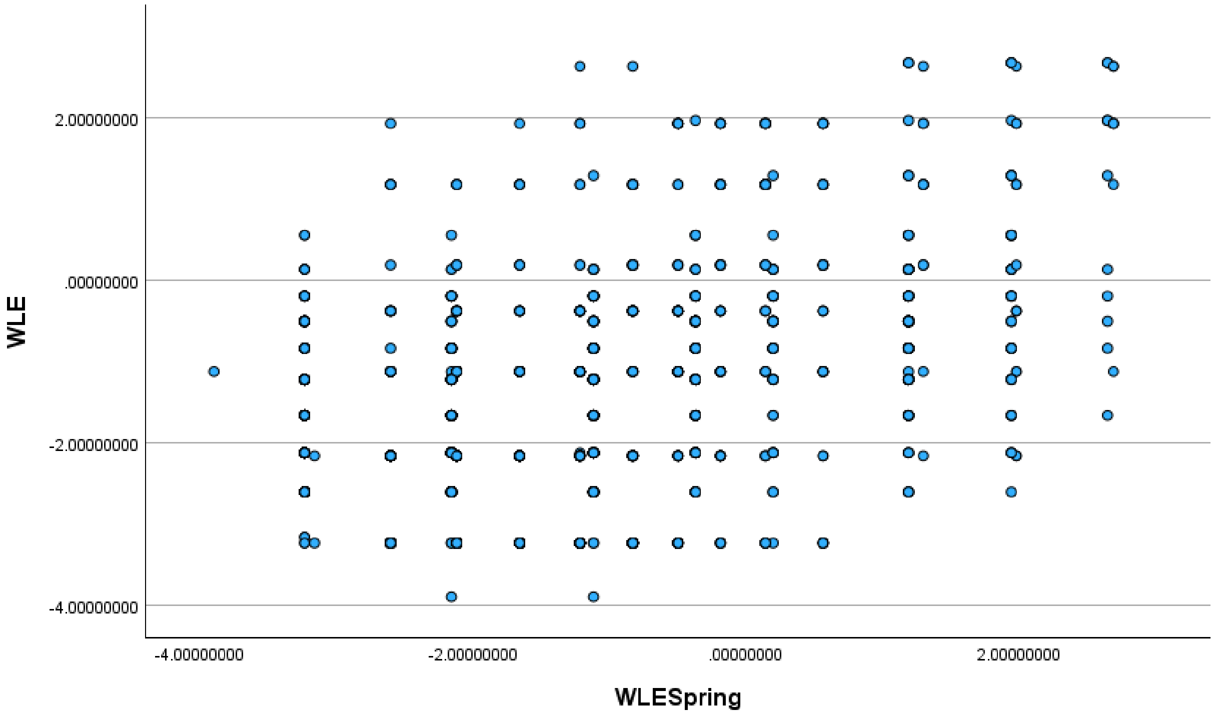

Figure 7. Cohort 3 Correlations between Fall and Spring DDM Scores