2026



STEM Action Center's Math Mentors Program: 2025-2026 Evaluation Report

Prepared for the
STEM Action Center
October 2026

THE UNIVERSITY OF UTAH
UTAH EDUCATION POLICY CENTER

THE UNIVERSITY OF UTAH

UTAH EDUCATION POLICY CENTER

Bridging Research, Policy, & Practice

The Utah Education Policy Center (UEPC) is an independent, non-partisan, not-for-profit research-based center at the University of Utah founded in the Department of Educational Leadership and Policy in 1990 and administered through the College of Education since 2007. The UEPC's mission is to bridge research, policy, and practice by conducting rigorous and comprehensive research and evaluations, and providing expert and research-informed technical assistance and professional learning. We empower educators, policymakers, and leaders to make research actionable and impactful to transform education across early childhood education, K-12 schools, and higher education.

We are committed to supporting the understanding of whether educationalpolicies, programs, and practices are being implemented as intended, whether they are effective and impactful, and how they may be improved and scaled-up and become sustainable.

Please visit our website for more information about the UEPC:
http://uepc.utah.edu

Andrea K. Rorrer, Ph.D., Director
andrea.rorrer@utah.edu

Cori Groth, Ph.D., Associate Director
cori.groth@utah.edu

Suggested Citation: Reynolds, A. L., Altermatt, E. R., Yildiz, M., Morris, S. R., Rorrer, A. K. & Wisham, L. (2026). *STEM Action Center's Math Mentors Program: 2025-2026 Evaluation Report.* Salt Lake City, UT: Utah Education Policy Center.

Copyright 2026, all rights reserved. ©

Acknowledgements

The Utah Education Policy Center thanks Clarence Ames, Jana Alexander, and Tami Goetz, all of the Utah STEM AC, for their leadership and dedication to supporting administrators, teachers, and students through the Math Mentors Program. We also appreciate the cooperation of school district and charter school administrators who supported evaluation efforts and facilitated educator and student participation. We are especially grateful to the teachers, mentors, and students who took the time to complete surveys and share their experiences in focus groups.

Table of Contents

Executive Summary *8*

Overview **8**

Data and Methods **8**

Key Findings **9**

Recommendations **10**

Introduction *11*

Evaluation Overview **11**

Summary of Key Evaluation Findings and Recommendations for Prior Year **11**

Evaluation Questions **12**

Report Organization **12**

Previous Interim Reporting **13**

Background *13*

Effectiveness of High-Dosage Tutoring **14**

Stacking High-Dosage Tutoring with Math Learning Software **14**

Challenges to Scaling High-Dosage Tutoring **15**

Evaluation Design and Data Sources *15*

Data Sources and Methods **16**

Data Sources 16

Methods 20

Implementation Evaluation Findings *20*

STEM AC Math Mentors Program Model **20**

Program Progress and Implementation **21**

Overview of Program Growth and Progress Since Year Three 21

Staffing and Program Leadership 23

Recruiting and Placing Mentors 23

Schools and Communities Served 23

Students Reached 27

Tutoring Dosage and Intensity **27**

Small-Group Tutoring Was Consistent, but MLS Use and Teacher Communication Were Uneven 30

Mentors Rated Training and Support Positively and Identified Areas to Improve **31**

Mentor Training 31

Ongoing Support for and Monitoring of Mentoring 36

Outcome Evaluation Findings: Change in Student Attitudes *38*

Student Survey Respondents **38**

Measures 39

Changes in Students' Mathematics Attitudes **39**
Mentor Success Story: "I Almost Like Math Now" **42**
Students' Experiences with Mentors and Mentoring and Math Learning Software **42**
Examining Differences by Grade Level **43**
Outcome Evaluation Findings: Student Achievement ***45***
Analytic Approach for Student Achievement Outcomes **46**
Measures 46
Data Analysis Strategy 47
Results from Descriptive Analyses **48**
Participation Rates and Median Tutoring Time by LEA 48
Math Achievement Outcomes by Participation Status 48
Math Achievement Outcomes by Tutoring Dosage 50
Results from Model-Based Analyses **51**
Adjusted Association Between MM Participation and Math Achievement 51
Adjusted Association Between Tutoring Dosage and Math Achievement 52
Sensitivity Analyses 53
Conclusions and Limitations for Student Achievement Findings **54**
Recommendations ***56***
Continue to focus on consistently reaching high-impact tutoring benchmarks for all students **56**
Strengthen the instructional integration of tutoring and use of MLS **57**
Build a shared understanding of and expectations for the mentor role **57**
Develop mentors' math tutoring practice through targeted support **58**
Strengthen program data infrastructure for monitoring, support, and evaluation **59**
References ***60***
Appendices ***64***
Appendix A. Mentor Focus Groups Report (May 2026) **64**
Appendix B. Interim evaluation report: Retrospective analysis of outcomes for the STEM Action Center's 2024-25 Math Mentors Program (February 2026) **69**
Appendix C. Supplementary tables for student surveys **76**
C1. Scale reliabilities for student survey measures 76
C2. Scale and item-level student survey results summary 76
C3. Grade-level regression models for student survey outcomes 77
Appendix D. School staff survey results summary **78**
Appendix E. Sensitivity analyses for achievement outcomes **80**
Project Staff ***81***

List of Tables

Table 1. Data Sources by Evaluation Question 15
Table 2. STEM AC MM Participation in Data Collection by Source and LEA 16
Table 3. Characteristics of Schools Participating in STEM AC MM by LEA 26
Table 4. Tutoring Dosage by LEA 28
Table 5. Participation Rates and Median Tutoring Time by LEA 48
Table 6. Adjusted Association Between MM Participation and End-of-Year Math Achievement, by LEA 52
Table 7. Adjusted Association Between MM Tutoring Dosage and End-of-Year Math Achievement, by LEA 52

List of Figures

Figure 1. Summary of Y4 STEM AC MM Program Goals 21
Figure 2. Growth in Program Reach and Tutoring Dosage, Year 3 to Year 4 22
Figure 3. Geographic Spread of Schools Participating in STEM AC MM, 2025-2026 24
Figure 4. Math Proficiency of Participating LEAs and Schools Relative to the Statewide Average 25
Figure 5. The Relationship Between the Concentration of Students Identified as Economically Disadvantaged and Math Proficiency Rates Across Participating Schools 26
Figure 6. STEM AC MM Program Reach in 2025-26 27
Figure 7. Distribution of Tutoring Dosage in Year 3 (2024-2025) as Compared to Year 4 (2025-2026) 28
Figure 8. Mentors' End-of-Session Ratings of the Trainings 31
Figure 9. Mean Ratings of Mentors' Understanding of Topics Before and After Training 32
Figure 10. Perceptions of Mentoring and Tutoring Competencies 33
Figure 11. Mentors' Perceptions of Support 36
Figure 12. Mentor Overall Experience in the MM Program 37
Figure 13. Student Survey Respondents and Response Rates by LEA and Grade 38
Figure 14. Pre/Post Change in Students' Mathematics Attitudes 40
Figure 15. Mentors' Perceived Impact on Students 41
Figure 16. Students' Ratings of Mentoring and Math Learning Software 43
Figure 17. Predicted Ratings of Mentoring and Math Learning Software by Student Grade Level 44
Figure 18. Gap In Average Standardized Math Scores Between MM Participants and Non-Participants at Baseline and End of Year, by LEA 49
Figure 19. Gap in Average Standardized Math Scores Between Each Tutoring Dosage Group and Non-Participants at Baseline and End of Year, by LEA 50

List of Acronyms

- **CCID**—Center for Creativity, Innovation and Discovery
- **FPA**—Freedom Preparatory Academy of Provo
- **GPA**—Gateway Preparatory Academy
- **HDT**—high dosage tutoring
- **IVI**—interest, valuing, and identity
- **LEA**—Local Education Agency
- **MLS**—Math Learning Software
- **MM**—Math Mentors
- **NSSA**—National Student Support Accelerator
- **SD**—School district
- **STEM AC**—STEM Action Center
- **UEPC**—Utah Education Policy Center
- **UICS**—Utah International Charter School

Executive Summary

Overview

This executive summary presents key findings and recommendations from the Utah Education Policy Center's (UEPC) implementation and outcome evaluation of the STEM Action Center's (STEM AC) Math Mentors (MM) program for the 2025-2026 school year (Year 4). In the STEM AC MM program, trained mentors provide up to 90 minutes of in-school math tutoring per week to students individually or in small groups as they complete activities in a math learning software (MLS) program, an approach that pairs two interventions in a "stacked intervention." Year 4 was the first program year operated solely by the STEM AC using state program funding. From the pilot year through Year 3, the program operated as the federally funded AmeriCorps Math Mentors Program (AMMP). Methodological details and comprehensive findings are provided in the full report.

The Year 4 evaluation addressed four evaluation questions (EQs):

EQ1. **Implementation.** To what degree is Math Mentors being implemented in ways that align with evidence-based and practice-informed guidelines, including those outlined by the NSSA?

EQ2. **Mentor Training and Support.** To what degree do mentors report that training and support opportunities were effective?

EQ3. **Student Attitudes.** Is participation in Math Mentors associated with positive changes in students' mathematics attitudes?

EQ4. **Student Achievement.** Is participation in Math Mentors associated with positive changes in students' mathematics achievement?

Data and Methods

The evaluation used a mixed methods design, drawing on quantitative and qualitative evidence from students, mentors, school staff, and administrative records. Data sources included the following:

- **Tutoring participation logs** designed and managed by the UEPC, documenting 24,098 tutoring sessions delivered by 75 mentors to 3,109 students
- **Surveys**
    - **UEPC Mentor Training Exit Survey** (75 responses from 53 mentors)
    - **UEPC EOY Mentor Survey** (45 mentors, 60% response rate)
    - **UEPC EOY Student Survey**, a retrospective pre- and post-survey of mathematics attitudes completed by 700 students (22.5% of students served)
    - **UEPC EOY Staff Survey**, completed by 57 school staff
- **Mentor focus groups** with 48 mentors from seven of the nine participating LEAs in 6 focus groups

- **Student demographic and mathematics achievement data from three selected LEAs**, including 4,877 students, of whom 1,087 participated in MM
- **Program artifacts** and **publicly available USBE school and LEA data**

Quantitative data were analyzed using descriptive and inferential statistics, including regression models estimated separately for each LEA in the student math achievement analyses. Qualitative data were analyzed using deductive coding of focus group transcripts and pattern coding of open-ended survey responses. The student achievement analyses are preliminary and focus on three LEAs. In spring 2027, the UEPC will conduct a more comprehensive analysis across all participating LEAs using statewide USBE student records.

Key Findings

The Year 4 evaluation produced five key findings across the four evaluation questions.

1. **The MM program exceeded its Year 4 goals for mentors and students served and increased tutoring dosage, but most students received less tutoring than NSSA recommends (EQ1).** The program placed 75 mentors (goal of 36) and served 3,109 students (goal of 2,000) across 20 schools in nine LEAs. The share of students receiving at least 450 minutes of tutoring rose from 19.6% to 41.0%, but most students received less than the roughly 900 minutes recommended by the National Student Support Accelerator (NSSA, 2025), and median dosage ranged from 125 minutes in Weber SD to 1,380 minutes in Granite SD. Part of the increase may reflect more complete documentation in the first year of the UEPC-managed participation log.

2. **Small-group tutoring was consistent, but MLS use and communication with teachers were uneven (EQ1).** Mentors delivered 93% of documented sessions individually or in small groups. MLS use was documented in 52.9% of sessions and ranged from none to all sessions across mentors, and 58% of mentors reported communicating weekly with teachers about student performance.

3. **Mentors rated training and support positively but were less confident in math tutoring tasks than in relationship building (EQ2).** Mentors rated support from school staff most highly (4.64 on a five-point scale), and 97.8% described their overall experience as positive or very positive. Confidence was lowest for using MLS to support learning and for helping students facing the greatest challenges, and understanding of the mentor role varied across schools.

4. **Students reported gains in mathematics attitudes, especially confidence and growth mindset, but rated MLS less positively than mentoring (EQ3).** Students reported statistically significant gains across all four measured attitudes (0.38 to 0.51 points on a five-point scale; small to medium effects), and more than eight in ten reported that their mentor helped them improve their math skills and confidence. Students rated MLS lower than mentoring (3.92 compared with 4.37), and students in eighth grade and above rated both less positively than third graders. Because only participating students were surveyed, these changes cannot be attributed to the program.

5 **Preliminary analyses found positive associations between MM participation and mathematics achievement in two of three focus LEAs selected for promising levels of tutoring dosage (EQ4).** In all three LEAs, participants began with lower achievement than non-participants. After adjusting for baseline achievement and student characteristics, participants scored an estimated 0.28 standard deviations higher than comparable non-participants in LEA 1 and 0.40 standard deviations higher in LEA 2, while the LEA 3 analyses showed no evidence of a positive association. These patterns held across alternative analyses, but they represent associations and not causal effects, and the spring 2027 analysis will examine all participating LEAs.

Recommendations

This year's recommendations are grounded in the Year 4 findings and aligned with design principles that support integrating high-dosage tutoring as an intervention to support student learning.

- **Continue to focus on consistently reaching high-impact tutoring benchmarks for all students.** Establish feasible tutoring schedules, provide guidance on mentor caseload size, and work with site supervisors to understand what facilitates or impedes reaching high-dosage tutoring benchmarks.
- **Strengthen the instructional integration of tutoring and use of MLS.** Define the instructional purposes of MLS within tutoring sessions, determine why MLS use is limited, and ensure mentors and sites have access to MLS training and technical assistance, including support in identifying programs suited to their students' grade levels.
- **Build a shared understanding of and expectations for the mentor role.** Develop a brief, shared description of the mentor's instructional role, establish routines for mentor and teacher communication, and engage site leaders and teachers in setting expectations about protected tutoring time at the start of the year.
- **Develop mentors' math tutoring practice through targeted support.** Refine training and ongoing support on tutoring competencies, differentiate training content by grade span, and create opportunities for mentors to observe and learn from peers.
- **Strengthen program data infrastructure for monitoring, support, and evaluation.** Maintain the participation log and monthly reporting structure, include participation log training and regular feedback in onboarding, and work with sites where participation in data collection has been lower to address barriers.

Introduction

Evaluation Overview

The STEM Action Center (AC) launched the Math Mentors program during the 2021-2022 school year, and the Utah Education Policy Center serves as the evaluator for this program. Math Mentors (MM) builds upon the STEM AC's K-12 Math Personalized Learning Software Program[1] by pairing a math software intervention with high-dosage tutoring. From the pilot year through Year 3, the program operated as the federally funded AmeriCorps Math Mentors Program (AMMP). Program Year 4 (2025-2026) was the first program year operated solely by the STEM AC utilizing state program funding. The program is now referred to as the STEM AC Math Mentors (MM).

As designed, students in the Math Mentors (MM) program work with a trained mentor who provides up to 90 minutes of in-school math tutoring per week to small groups of students as they complete activities in a math learning software program. Over its five-year period from 2021-2022 (Pilot Year) to 2025-2026 (Year 4), key aims of the program included:

- expanding access to high-dosage tutoring for students in Utah,
- ensuring that mentors have the support they need to be effective tutors and to build strong relationships with students, and
- improving students' confidence, engagement, and achievement in mathematics.[2]

In the pilot year of the program (2021-2022), the Utah Education Policy Center (UEPC) supported the STEM AC in developing a program that aligned with evidence-based and practice-informed guidelines for tutoring and mentoring programs, including recommendations (Robinson et al., 2024) from the National Student Support Accelerator (NSSA, 2025) for "high-impact" tutoring programs (e.g., at least three weekly tutoring sessions, usually a minimum of 30 minutes per session, over 10 or more weeks).

Since Year 1 (2022-2023), the UEPC has conducted implementation and outcome evaluation of the Math Mentors program. Data for these evaluations has included student participation data collected by the STEM AC, mentors' responses to UEPC surveys and interview protocols, site supervisors' responses to UEPC surveys, students' responses to surveys, and demographic and achievement data collected from LEAs by the UEPC. Annual reports were provided to the STEM AC in August of each year (Altermatt, Groth, & Rorrer, 2022; Altermatt, Altermatt, Rorrer, Groth, & Timmer, 2023; Altermatt, Yildiz, & Rorrer, 2024; Altermatt, Yildiz, & Rorrer, 2025a).

Summary of Key Evaluation Findings and Recommendations for Prior Year

To provide context for Year 4 (2025-2026) evaluation efforts and results, we first summarize key findings from Year 3 (2024-2025). These include findings related to the number of participating mentors, mentors' perceptions of training and support, the number of participating students, tutoring dosage, and changes in participating students' math attitudes and achievement outcomes.

[1] https://stem.utah.gov/educators/funding/k-12-math-personalized-learning-software-grant/
[2] https://stem.utah.gov/mathmentors/

During Year 3, the STEM AC placed 51 mentors in 14 Local Education Agencies (LEAs), providing them with both pre-service training and in-service support. Among mentors who completed training exit surveys (n = 12) and end-of-year surveys (n = 15), most reported that the pre-service training was effective and indicated that they felt supported throughout the year. However, some mentors expressed a need for additional training and support opportunities through open-ended survey responses. According to a participation log maintained by the STEM AC, mentors provided tutoring to 2,127 students, with 41% receiving at least 180 minutes (3 hours) of tutoring over the course of the year. Student survey data (n = 219) indicated statistically significant improvements in students' self-reported math confidence, effort, and interest in math success. Evidence of impact on achievement outcomes was less clear. Analyses of student standardized test scores indicated that the program was reaching students who were struggling in mathematics, statistically significant improvements in math performance were not observed.

Key recommendations for the prior year included: (1) continuing efforts to ensure that evidence-based and program-prescribed recommendations for high-dosage tutoring are met for participating students, (2) increasing training and support opportunities for mentors to ensure consistent implementation of program practices, and (3) expanding mentor, site supervisor, and student participation in data collection efforts to better monitor implementation and outcomes.

Evaluation Questions

For Year 4 (2025-2026), the UEPC sought to build upon prior year evaluation efforts by addressing four evaluation questions (EQs).

- **EQ1. Implementation.** To what degree is Math Mentors being implemented in ways that align with evidence-based and practice-informed guidelines, including those outlined by the NSSA?
- **EQ2. Mentor Training and Support.** To what degree do mentors report that training and support opportunities were effective?
- **EQ3. Student Attitudes.** Is participation in Math Mentors associated with positive changes in students' mathematics attitudes?
- **EQ4. Student Achievement.** Is participation in Math Mentors associated with positive changes in students' mathematics achievement?

Report Organization

This report is organized into eight sections. The *Executive Summary* presents an overview of the study, key findings, and recommendations. This *Introduction* provides an overview of the MM program, summarizes findings and recommendations from the Year 3 evaluation, and presents the evaluation questions guiding Year 4. The *Background* section reviews research on the effectiveness of tutoring and MLS interventions for improving student outcomes in mathematics. The *Evaluation Design and Data Sources* section describes the data that inform this evaluation, maps each source to the evaluation questions it addresses, and summarizes participation in data collection activities.

Three sections present the findings for the evaluation questions. The *Implementation Evaluation Findings* section examines how the program was implemented in Year 4, including program reach, tutoring dosage, fidelity to the program model, and mentors' perceptions of training and support (EQ1 and EQ2). *Outcome Evaluation Findings: Changes in Student Attitudes* examines changes in participating students' mathematics attitudes (EQ3). *Outcome Evaluation Findings: Student Achievement* presents preliminary analyses of the association between program participation and mathematics achievement in three selected participating LEAs (EQ4).

The final section, *Recommendations*, offers actionable steps for ongoing program improvement to help STEM AC strengthen and sustain the MM program.

Previous Interim Reporting

Importantly, the UEPC provided interim reports to the STEM AC throughout the academic year to support continuous improvement. These reports included a report summarizing findings from a training exit survey completed by mentors, monthly reports comparing records of hours invoiced by mentors with hours recorded in the participation log, a report summarizing key themes from focus groups with mentors (see Appendix A), and a report summarizing findings from retrospective analyses of associations between program engagement and student outcomes in Year 3 (2024-2025) (see Appendix B).

Background

This section reviews the research behind the Math Mentors program. It begins with the national and local context of declines in student math achievement since the pandemic, and why recovery remains a priority in Utah. We then turn to the evidence for high-dosage tutoring as a means of addressing student math achievement. First, offering an overview of the effectiveness of high-dosage tutoring, followed by the potential of added support of a stacked intervention leveraging math learning software, and concluding with the challenges of scaling successful programs.



Math achievement remains a concern in education in the United States. Last year's annual UEPC Math Mentors evaluation report also reviewed the substantial mathematics learning losses since the COVID-19 pandemic and the national and Utah-specific recovery efforts since (Altermatt et al., 2025a). As of Spring 2025, Utah math proficiency remains well below pre-pandemic levels. Only 45.1% of students were proficient, down from 49.7% baseline in Spring 2016. Math proficiency is substantially lower among students identified as economically disadvantaged (26.9%), English language learners (16%), and students with disabilities (19.1%) (USBE, 2026). Closing this recovery gap will continue to require sustained, multi-year investments in interventions with strong evidence of effectiveness (Betebenner & Van Iwaarden, 2024; Morton et al., 2025). Here we address high-dosage tutoring, which is one of those interventions identified to support learning, particularly among students who may struggle with math.

Effectiveness of High-Dosage Tutoring

Our prior report also summarized the growing body of evidence supporting high-dosage tutoring as an effective strategy for addressing unfinished learning (Altermatt et al., 2025a). This summary included the tutoring design features associated with the strongest results, such as frequent sessions embedded in the school day, low tutor-to-student ratios, consistent tutor-student pairing, and formative use of assessment data to inform instruction (Guryan et al., 2023; Nickow et al., 2020).

Math-focused tutoring shows meaningful effects across grade levels, including in middle and high school, where it is often harder to move the needle on student test-based outcomes (Guryan et al., 2023; Kraft et al., 2024a). For example, a randomized controlled trial (RCT) of high-dosage math tutoring in the Netherlands, including over 400 elementary students, found a 0.28 SD improvement, closing roughly 40% of the income-achievement gap after one year (de Ree et al., 2023). This differs from literacy interventions, whose benefits are often concentrated in early grades and diminish by upper-elementary years (Markovitz et al., 2022; Nickow et al., 2020; What Works Clearinghouse, 2017).

Stacking High-Dosage Tutoring with Math Learning Software

The STEM AC's Math Mentors program strategically pairs tutoring with math learning software (MLS), building on the idea that two interventions can be mutually reinforcing in a "stacked intervention" approach. This approach is supported by recent evidence, including a large multi-district study that found educator- and tutoring produced meaningful math gains while stand-alone digital-learning programs did not, supporting the case for pairing the two rather than relying on either alone (Carbonari et al., 2026). Human-delivered tutoring can support an increase in students' engagement with and productive use of educational technology, and technology helps tutors track student progress and individualize instruction (Bhatt et al., 2024; Cortes et al., 2025; Guryan & Ludwig, 2023; Robinson et al., 2025).

The 2025 UEPC evaluation report reviewed the broader evidence base for technology-enabled math instruction (Altermatt et al., 2025a; Altermatt et al., 2022; Hillmayr et al., 2020). Recent peer-reviewed meta-analyses reinforce and extend this evidence. A 2025 meta-analysis of 21 studies (40 samples) found AI-based math tools produced a small overall positive effect on K-12 achievement ($g = 0.342$) (Yi et al., 2025). Two types drove this effect: intelligent tutoring systems, which give students step-by-step feedback ($g = 0.432$), and adaptive learning systems, which adjust content to student performance ($g = 0.347$), while software using on-screen pedagogical agents showed no significant effect ($g = 0.046$). Another 2026 meta-analysis of AI-supported instruction across subjects found a similar moderate effect on student learning outcomes (Liu et al., 2026). But other evidence remains mixed. For example, in one large field trial embedded in a widely used adaptive math program, AI-revised word problems reduced errors on simpler two-step equation problems but increased errors on more complex problems involving proportional relationships and graphing linear relationships (Norberg et al., 2025). Together with continuing evidence that engagement and usage levels moderate software's impact (Di Pietro & Munoz, 2025; Eames et al., 2026), the current research landscape supports the STEM AC Math Mentors model's combination of consistent human-led tutoring with structured MLS use.

Challenges to Scaling High-Dosage Tutoring

A central challenge identified in high-dosage tutoring literature remains whether scaling high-dosage tutoring dilutes its effectiveness. As programs grow, dosage tends to decline, tutor-to-student ratios tend to rise, and implementation quality tends to suffer, sometimes eliminating measurable effects entirely (Kraft et al., 2024a; Kraft et al., 2024b). A 3-year RCT of a university-led high school tutoring program in Oklahoma, including over 500 ninth-grade students, found a smaller but positive effect (0.13 SD) at a cost of about $4,740 per student (Hamlin et al., 2026). The program also found no drop in effectiveness when moving from a 2:1 to a 3:1 tutor ratio, suggesting a possible way to reduce costs while scaling up programs.

Additionally, a 2026 study by Shmoys and colleagues of five tutoring programs in math and reading found that none reached the field's high-dosage threshold of three or more sessions per week for at least 10 weeks, delivering under two-thirds of planned hours due to scheduling, staffing, and technology problems. One emerging response to this challenge is pairing in-school math tutoring with educational technology: a multi-district initiative using this model has maintained tutoring quality and dosage while scaling from 9 to 18 schools over its first 2 years of implementation (University of Chicago Education Lab, 2025).

Evaluation Design and Data Sources

The UEPC evaluation of the 2025-26 STEM AC MM program year used a mixed methods design, drawing on quantitative and qualitative evidence from students, mentors, school staff, and administrative records. We organize findings by topic, integrating evidence from multiple sources. Table 1. Data Sources by Evaluation QuestionTable 1 maps each data source to the evaluation questions it informs. As noted in Table 1, no evaluation question rests on a single source, and three of the four draw on both quantitative and qualitative evidence. The following sections describe each source in more detail.

Table 1. Data Sources by Evaluation Question

	EQ1 – Implementation	EQ2 – Mentor Training & Support	EQ3 – Student Attitudes	EQ4 – Student Achievement
Program Artifacts	✓			
Participation logs	✓			✓
Mentor Training Exit Survey		✓		
EOY Mentor Survey	✓	✓	✓	
Mentor focus groups	✓	✓	✓	

	EQ1 – Implementation	EQ2 – Mentor Training & Support	EQ3 – Student Attitudes	EQ4 – Student Achievement
EOY Student Survey			✓	
EOY School Staff Survey	✓	✓		
LEA student achievement and demographic data				✓
Supplementary data (e.g., state report cards)	✓			

Data Sources and Methods

The Year 4 evaluation drew on data collected by the UEPC as well as materials generated by the STEM AC during the course of operating the program. Each source is described below. Table 2 summarizes participation across sources and LEAs. We then broadly discuss the methods used in the evaluation, while detailed approaches to analyses are embedded within the findings.

Table 2. STEM AC MM Participation in Data Collection by Source and LEA

LEA	Participation Logs		Surveys				Focus Groups
	# students	# mentors	Mentor training survey	EOY Mentor Survey	EOY Student Survey	EOY School Staff Survey	Mentor focus groups
Cache	1,021	30	25	18	264	38	25
CCID	235	7	3	4	12	0	3
FPA	249	3	3	1	1	1	2
GPA	312	9	6	8	57	1	6
Granite	113	5	3	2	4	3	5
Millard	293	5	0	1	52	4	0
San Juan	267	7	3	1	28	0	0
UICS	129	3	3	3	28	0	2
Weber	490	6	6	4	245	4	5
Unknown	0	0	1	3	9	6	0
Total	**3,109**	**75**	**53**	**45**	**700**	**57**	**48**

Data Sources

The UEPC team drew on nine data sources to inform the evaluation. These included both qualitative (e.g., program artifacts, focus groups, open-ended survey responses) and quantitative data sources (e.g., surveys, participation logs, LEA achievement and demographic data). The following sections

detail each data source, including descriptions of data collection and analysis, where relevant, further details of analyses are included in the results sections.

Program Artifacts

Several materials from the STEM AC MM team informed this evaluation. The program's logic model and strategic goals, developed during the pilot year (2021-2022), establish the design and targets against which Year 4 implementation is assessed. Ongoing communications between the UEPC and STEM AC MM program leaders, including regular meetings and correspondence about program operations, provided context on staffing, program changes, and implementation conditions encountered during the year. In addition, mentors voluntarily submitted brief success stories to STEM AC MM program leaders this year. These were not collected as evaluation data and are used illustratively rather than analytically, with identifying information removed.

Participation Logs

Mentors documented each tutoring session in a centralized log designed and maintained by the UEPC, recording session date, students served, session length, group size, use of math learning software, and session notes. This was the first program year that the UEPC team managed and monitored the tutoring participation logs. The UEPC monitored submissions monthly, resolved inconsistencies in student and mentor identifiers, and produced interim reports for the STEM AC MM team that summarized invoiced versus logged hours for active mentors and flagged issues that warranted the team's attention throughout the year to improve data quality. Participation logs capture only the tutoring sessions mentors recorded and do not document the full range of activities for which mentors are compensated. The participation logs included 24,098 entries from the 75 mentors in Year 4.

Because tutoring dosage is measured from the participation logs, it is important to be clear about what those logs do and do not capture in terms of mentors' time. Mentors record the sessions they conduct with students. They do not record the full range of activities for which they are compensated, which also includes program and district trainings, meetings with teachers and site supervisors, preparation, and the time required to log the tutoring sessions in the participation logs.

Three factors account for the observed gap between documented tutoring and invoiced hours in the past program year, each of which has different implications.

- **Time allocated to non-tutoring responsibilities.** Training, collaboration with school staff, and administrative tasks are necessary components of the mentor role. Time spent on them is a part of the requirements of the role and should not be counted as tutoring time, which would diminish the accuracy of the measure for estimating program impacts on student outcomes.
- **Tutoring that occurred but was not documented.** Mentors report that sessions are occasionally not logged, particularly when they move directly between sessions without time to record, or when work with students takes a form they do not recognize as a loggable session, such as math games or whole-class support. This suggested that the dosage documented in the participation logs may undercount total minutes of tutoring.

- **Scheduled tutoring time that was lost.** Some tutoring that was scheduled did not occur. In correspondence with the UEPC, STEM AC staff identified a range of interruptions reported by mentors, including student absences, field trips, assemblies, school drills, students being pulled for competing interventions, students arriving late to sessions, and instances in which teachers indicated that students did not need to be pulled on a given day. These logistical issues may contribute barriers to achieving the recommended tutoring minutes during the year.

As a consequence of better understanding discrepancies between invoiced and logged tutoring hours, the STEM AC and UEPC teams discussed how to determine a realistic target for the ratio of documented tutoring hours and invoiced hours. It was determined that the initial target of 75% was unlikely to be attainable, given the responsibilities described above, and it was suggested that roughly 50% represented a more realistic expectation.

Mentor Training Exit Survey

Mentors were invited to complete a brief UEPC Mentor Training Exit Survey after completing each required training (e.g., STEM AC and Saga Coach). The survey included three sets of Likert-scale items assessing mentors' overall perceptions of the training, self-reported changes in their understanding of the topics covered, and confidence in the competencies necessary to be effective tutors and mentors. Two open-ended items asked mentors to identify the most valuable aspects of the training and areas for improvement. Between the end of August 2025 and mid-May 2026, when data collection ended, the UEPC received 75 completed responses from 53 unique mentors. Because mentors completed the survey after each training they attended, some mentors are represented twice. The median response time was 3.4 minutes. An interim report summarizing responses received through November 1, 2025, was provided to the STEM AC to inform ongoing implementation.

EOY Mentor Survey

Mentors were invited to complete the UEPC EOY Mentor Survey in spring 2026. The survey included four sets of Likert-scale items assessing how consistently mentors engaged in core program practices, their perceptions of the onboarding, training, and support they received, their perceptions of the program's impact on the students they worked with, and their overall experience in the program. Open-ended items asked mentors to describe what made the program work well for students, how the program might be improved, and how participating in the program had affected them. Forty-five mentors completed the survey, a 60% response rate, with respondents from all nine participating LEAs.

Mentor Focus Groups

The UEPC conducted six virtual focus groups with 48 mentors in March 2026, reaching mentors from seven of the nine participating LEAs and 14 of the 20 participating schools. Focus group sessions followed a semi-structured protocol, lasted an average of 44 minutes, and were recorded and transcribed. Transcripts were coded deductively to identify recurring themes. Focus group analysis was conducted by the UEPC evaluation team (see Appendix A for a full summary).

EOY Student Survey

Students who had received tutoring through the STEM AC MM program in third grade or above were invited to complete an anonymous retrospective pre- and post-survey assessing changes in their

mathematics attitudes over the course of their participation in the program (i.e., UEPC EOY Student Survey).[3] This within-subjects survey design asked students to rate each item twice, once for how they felt before working with their mentor and once for how they felt at the time of the survey. This design reduces respondent burden, permits paired analysis of anonymous responses, avoids the response shift bias that can affect conventional pre/post designs when students' understanding of a construct changes over the course of a program, and has been demonstrated as an effective approach to assess outcomes in youth serving programs (e.g., Little et al., 2020; Pratt et al., 2000; Young & Kallemeyn, 2019).

A total of 700 students completed the survey, representing 22.5% of the 3,109 students served and more than three times the 219 responses received in the Year 3 student survey. We received responses from students from all nine participating LEAs and across grade levels eligible to participate in the survey.

EOY Staff Survey

School staff at participating sites were invited to complete the UEPC EOY Staff Survey in spring 2026. The survey included three sets of Likert-scale items assessing how consistently staff observed mentors engaging in core program practices, their perceptions of the MM program's impact on participating students, and their overall experience with the MM program. Open-ended items asked staff to describe what made the program work well for students and how it might be improved. Fifty-seven school staff completed the survey, with respondents representing six of the nine participating LEAs. Because the number of eligible staff at each site is not known, a response rate cannot be calculated. Two-thirds of responses came from a single LEA. Given this uneven coverage, quantitative results from the staff survey are reported separately in Appendix D, and open-ended responses are used selectively to examine whether staff describe patterns consistent with what we found in other data sources.

LEA Achievement and Demographic Data

Student demographic and mathematics achievement data were collected from three participating LEAs.[4] These LEAs were purposively selected based on relatively high numbers of student participants and tutoring dosage, supporting estimation of associations between program engagement and student outcomes under conditions of meaningful implementation. Together, they account for just over one-third of all MM participants in Year 4. Subsequent retrospective analyses will use statewide

[3] The UEPC EOY Student Survey is administered under the Data Sharing Agreements between the UEPC and participating LEAs. Parents and guardians receive a letter of information with the option to opt their student out, and students may decline to participate or skip any question. The surveys do not contain sensitive questions that would require parent/guardian consent. Results are reported only in aggregate and do not include any identifiable data, consistent with the agreements, the UEPC's IRB approval, and its data sharing agreement with USBE.

[4] Data Privacy Agreements were established between the UEPC and LEAs participating in the evaluation of the STEM AC Math Mentors program. In addition to adhering to the DPA stipulations, the UEPC maintains University of Utah Institutional Review Board approval for conducting research with human subjects. Only authorized UEPC staff may access the data and data are used only for the purposes for which they were collected.

USBE records to assess generalizability across all participating LEAs. Student mathematics achievement was measured with Acadience Math for K–3 and RISE Math for grades 4–6.[5]

Supplementary Data

Supplementary school- and LEA-level data were also accessed via the publicly available USBE State Report Cards and Data Gateway (2025-26). These data supported the generation of descriptive statistics for the school and LEA contexts in which the STEM AC MM program operates.

Methods

This UEPC evaluation uses a mixed methods approach, drawing on the quantitative and qualitative data sources described above. We analyzed quantitative data using a combination of descriptive and inferential statistics. We analyzed qualitative data using deductive coding of focus group transcripts and pattern coding of open-ended survey responses and session notes to identify themes (Saldaña, 2025). Throughout the report, we integrate findings across sources to more fully address the evaluation questions. Additional detail on analytic decisions appears within the relevant findings sections, where those decisions inform how results should be interpreted in context.

Implementation Evaluation Findings

In this section of the report, we provide an implementation evaluation of the Math Mentors program for Year 4 (2025-2026), with the goal of answering the first two evaluation questions:

> **EQ1.** To what degree is Math Mentors being implemented with fidelity and in ways that align with evidence-based and practice-informed guidelines for high-dosage tutoring?
>
> **EQ2.** To what degree do mentors report that training and support opportunities were effective?

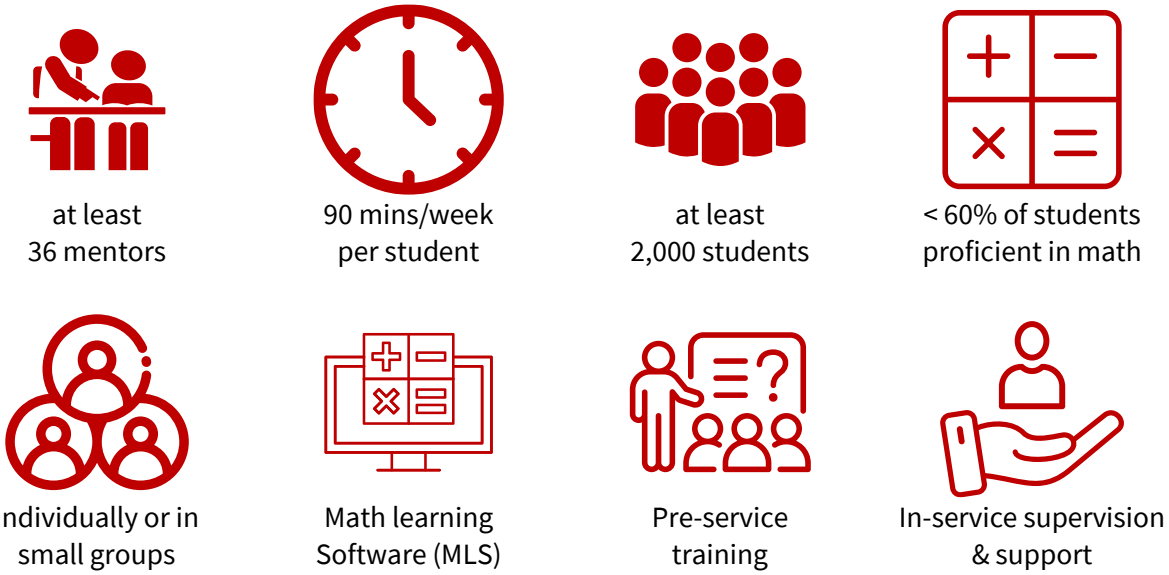
STEM AC Math Mentors Program Model

During the 2021-2022 pilot year, STEM AC personnel developed a logic model and strategic plan for Math Mentors, both of which are included in the Pilot Year report (Altermatt et al., 2022). Figure 1 reflects current program goals, including recruiting 36 mentors to collectively provide 90 minutes of in-school math tutoring per week to 2,000 students attending LEAs where fewer than 60% of students are grade-level proficient in math. These program goals reflect an increase from Year 3 when the program aimed to recruit 22 mentors and provide tutoring for at least 1,410 students. Consistent with recommendations from the National Student Support Accelerator (NSSA), tutoring was designed to be delivered individually or in small groups and supported by high-quality instructional materials, such

[5] The Utah Education Policy Center has a Master Data Sharing Agreement with the Utah State Board of Education for use of education data for evaluation and research purposes. The UEPC adheres to terms of the Master Data Sharing Agreement, including terms of use, confidentiality and non-disclosure, data security, monitoring, and applicable state and federal laws, and the University of Utah Institutional Review Board provisions, for the purposes of educational evaluations and research. The UEPC also adheres to FERPA regulations to protect student privacy. Consistent with FERPA and data share agreements, the UEPC does not report data for samples with fewer than 10 students. In addition, the UEPC does not report any personally identifiable information.

as evidence-based math learning software. Mentors were to be provided with both pre-service training and ongoing supervision and support.

Figure 1. Summary of Y4 STEM AC MM Program Goals



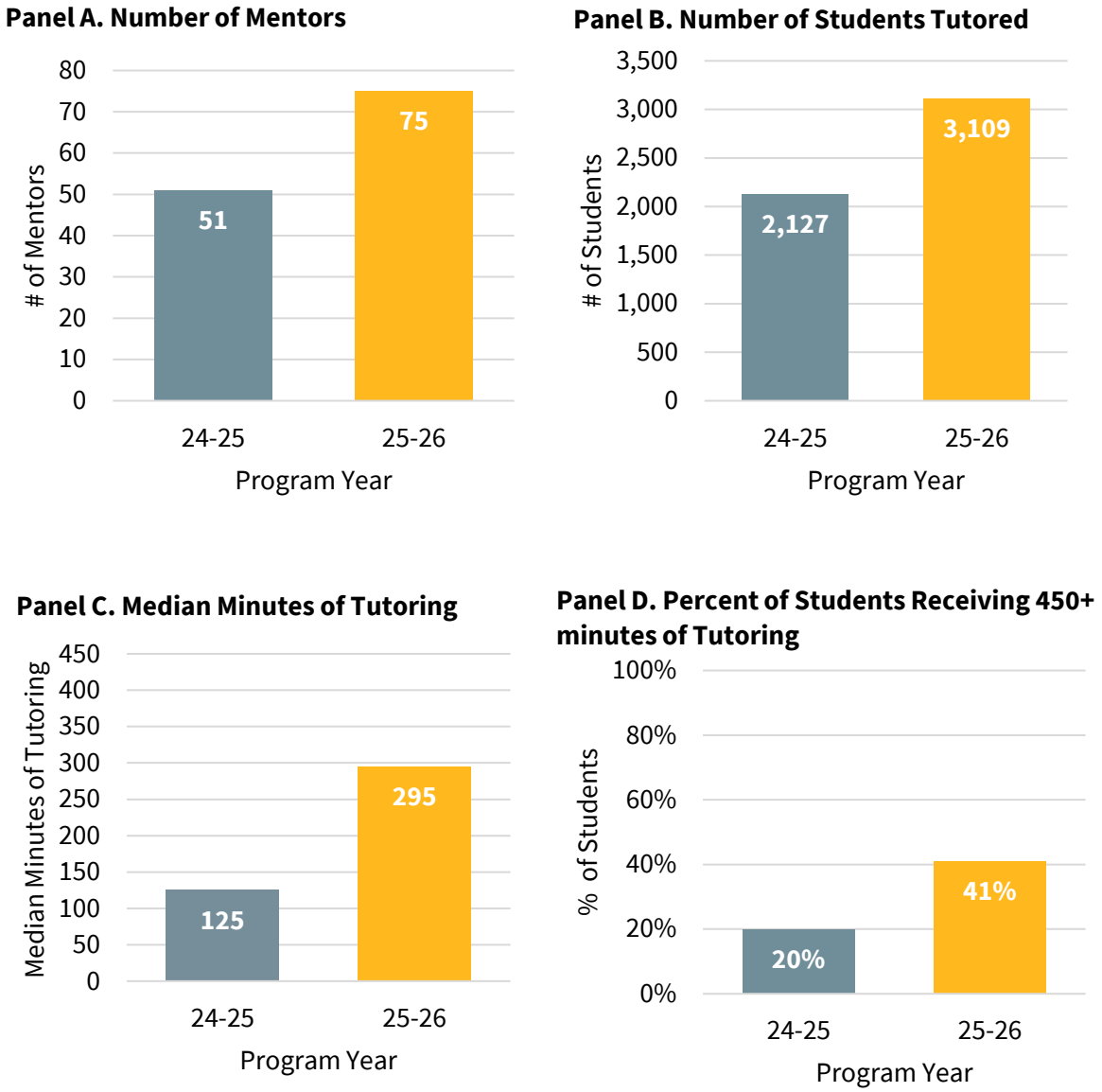
Program Progress and Implementation

The following sections describe how the STEM AC MM program has progressed since Year 3 program implementation in the current year. We begin with an overview of notable areas of program growth and progress. This is followed by changes in program staffing and leadership, most notably that this was the first program year in which the STEM AC fully operated the program without AmeriCorps fiscal support. It then covers how mentors were recruited and placed, the schools and communities where they served, and the students they reached. The sections that follow build on this picture of who delivered the program and where it was delivered.

Overview of Program Growth and Progress Since Year Three

Year 4 was a year of growth in both the STEM AC MM program's scale and intensity. The program recruited and trained more mentors, served more students, delivered substantially more tutoring to each student, and did so with a program that was more consistently documented than in the prior year. Figure 2 summarizes the change across four indicators.

Figure 2. Growth in Program Reach and Tutoring Dosage, Year 3 to Year 4



The program grew from 51 to 75 mentors and from 2,127 to 3,109 students served, meeting its goals for mentors and students served for 2025-26. The amount of tutoring each student received also meaningfully grew. Median dosage more than doubled, from 125 to 295 minutes, and the share of students receiving at least 450 minutes of tutoring rose from 19.6% to 41.0%, a 21-percentage-point shift examined in detail later in this section. These gains followed priorities set by the STEM AC MM team set in response to the Year 3 evaluation findings.

Two changes in Year 4 also improved the quality of the evidence available for this evaluation. First, the UEPC designed and actively managed the participation log for the first time, monitoring submissions on a monthly basis, routinely resolving inconsistencies in student and mentor identifiers, and reporting on tutoring logged back to the STEM AC and to site supervisors throughout the year. Second, as noted in the 2024-2025 evaluation, recommendations were made to increase participation in evaluation activities. Participation rose substantially across all instruments. Both changes mean

that Year 4 offers a rich picture of the program. The changes in available evidence also warrant a note of caution in interpreting year-over-year comparisons, as an unknown portion of the increase in documented tutoring may reflect simply increased documentation in the participation logs instead of additional tutoring delivered.

Staffing and Program Leadership

In Year 4, the STEM AC Math Mentors program was administered under the leadership of a two-person team from the STEM AC. With expertise in STEM education, educational technology, and grants and fiscal management, this team oversaw all aspects of the program, including recruitment, onboarding, training, supervision, and ongoing support. This was also the first program year that the STEM AC fully funded the program, no longer receiving fiscal support through the federally funded AmeriCorps program as was the case in previous years.

To further strengthen training and support, a veteran mentor was identified in Year 4 to serve in a lead mentor role. This mentor developed monthly newsletters, served as an additional point of contact for mentors with questions, and provided regional training and support to newer mentors. As discussed later, mentors' accounts of the supports available to them suggest that this addition addressed a need identified in the prior year.

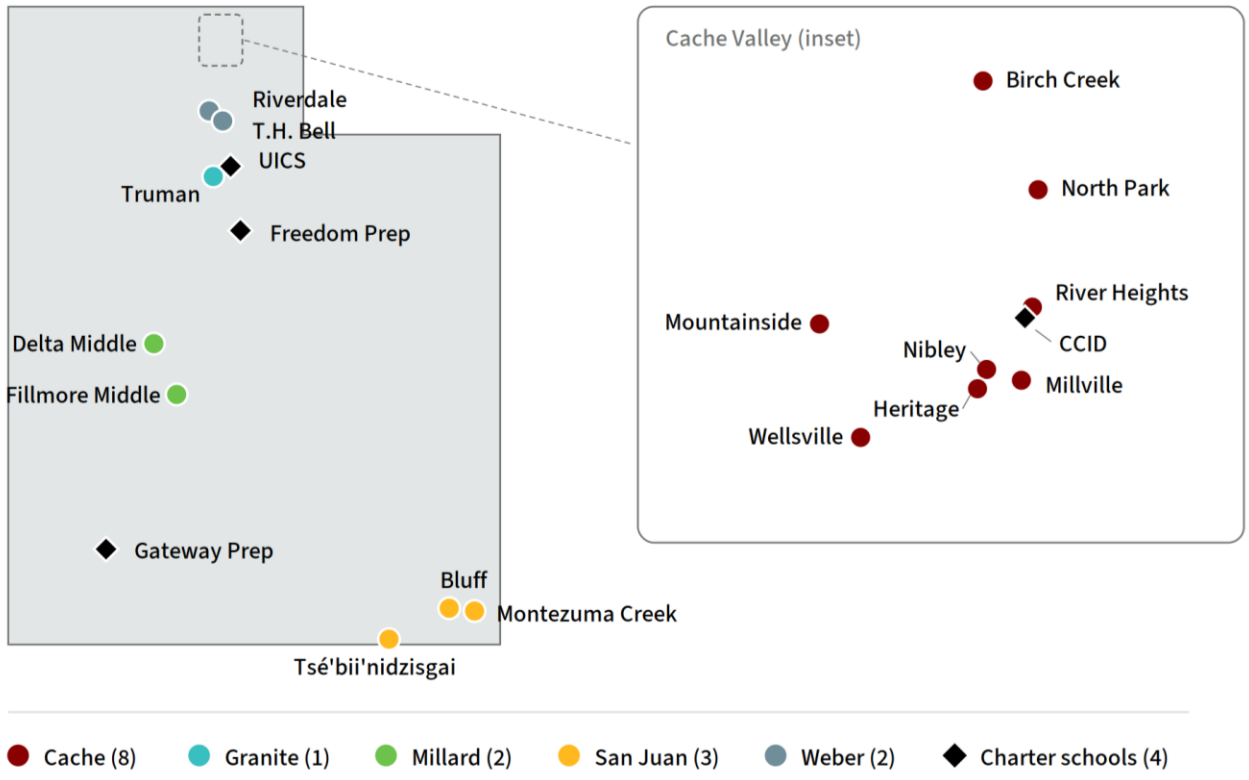
Recruiting and Placing Mentors

STEM AC staff exceeded the goal of recruiting, onboarding, and placing at least 36 mentors. In all, 75 mentors served in Math Mentors during Year 4, up from 51 in Year 3. Recruitment occurred on a rolling basis throughout the school year. Thirty-nine of the 75 mentors returned after serving in one or more prior program years, and 19 began their service after December 31, 2025. The high return rate of mentors is notable. Retaining more than half of the mentor corps from one year to the next means that a substantial share of tutoring in Year 4 was delivered by mentors who already knew the program’s expectations. This report does not examine whether returning mentors stayed at the same schools or worked with the same students.

Schools and Communities Served

Mentors were placed in 9 LEAs in Year 4, down from 14 in Year 3, and served 20 schools in total. These schools included 16 elementary and middle or junior high schools across five public school districts (Cache County School District; Granite School District; Millard School District; San Juan School District; and Weber School District) and four charter schools (Center for Creativity, Innovation, and Discovery [CCID]; Freedom Preparatory Academy of Provo [FPA]; Gateway Preparatory Academy [GPA]; and Utah International Charter School [UICS]). The drop from 14 to 9 LEAs means mentors were concentrated at fewer sites, while the program served roughly 1,000 more students in Year 4 than in Year 3. Participating schools are geographically distributed across the state, including areas off the Wasatch Front, as shown in Figure 3.

Figure 3. Geographic Spread of Schools Participating in STEM AC MM, 2025-2026

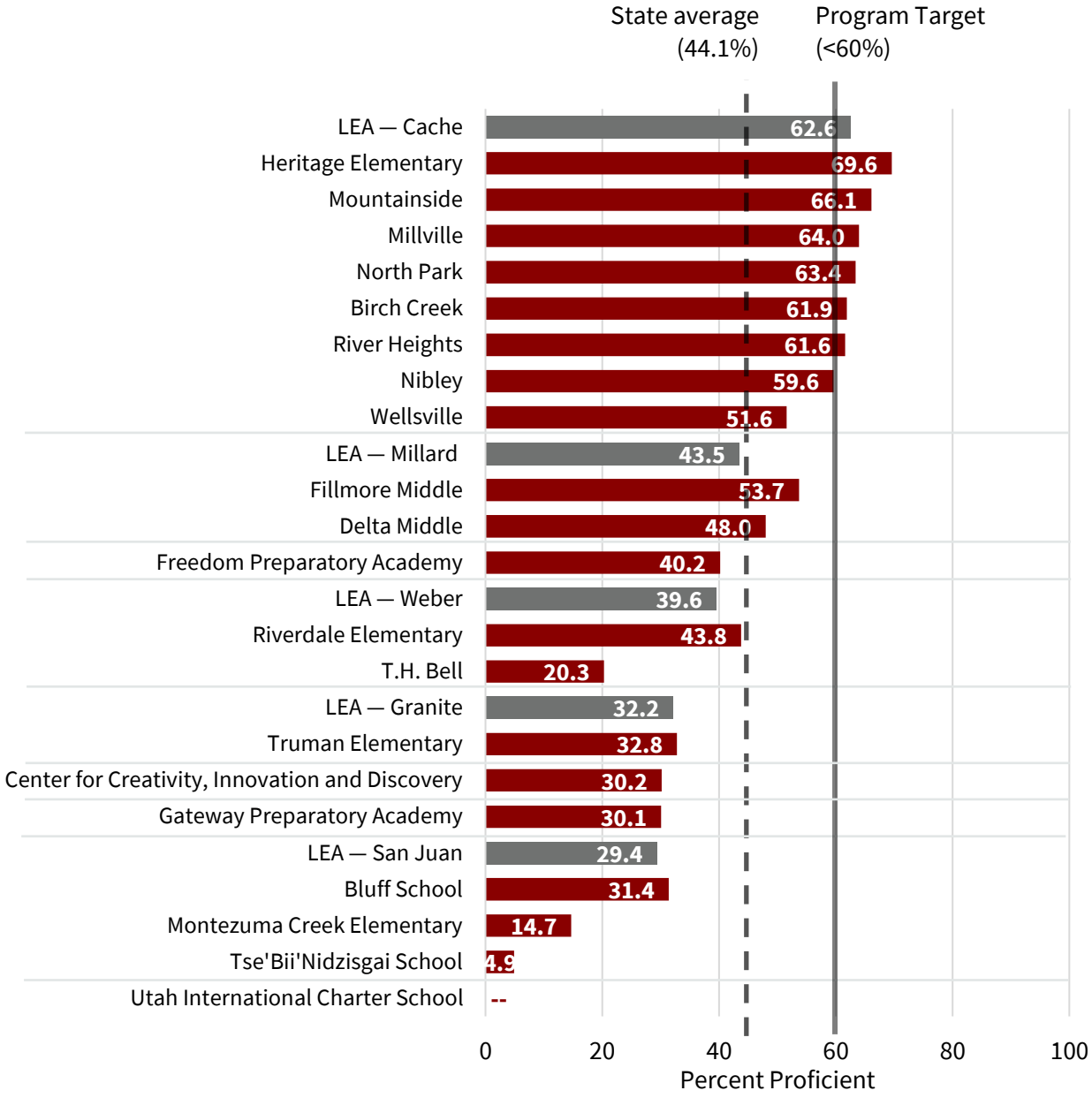


Source: STEM AC MM program data

We note that although the STEM AC MM program reaches schools across Utah, nine of the 20 schools located are within the Cache Valley, and the Cache School District alone accounts for approximately one-third of all students served by the MM program in 25-26 (see Table 3 for school characteristics by LEA). This variation in context speaks to how the program is reaching a range of school communities across the state.

Consistent with the goal of placing mentors in schools and districts in which fewer than 60% of students are proficient in mathematics, the program continued to serve schools with substantial academic need. As shown in Figure 4, the percentage of students proficient in math was below 60% for eight of the nine participating LEAs for which data were publicly available.

Figure 4. Math Proficiency of Participating LEAs and Schools Relative to the Statewide Average

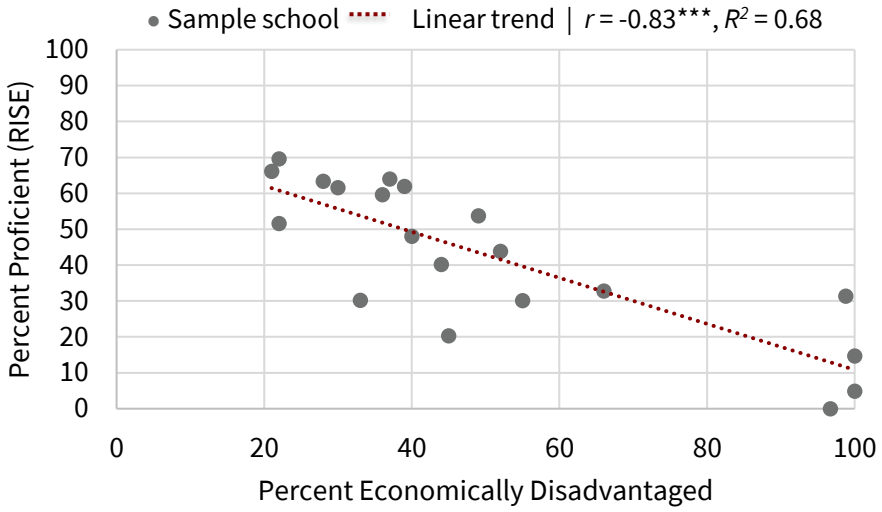


Note. Grey bars indicate LEA average percent proficient on RISE math in 2024-25, including all schools, red bars indicate the percent proficient for sample schools. The vertical dashed line indicates the state percent proficient (44.1%) and the solid vertical line indicates the program target of serving schools with <60% of students proficient in math. Proficiency data for UICS were suppressed.
Sources: USBE 2025-26 Report Cards

As shown in the Figure 4, Cache SD is the exception, with a districtwide proficiency rate above the MM program's less than 60% proficiency threshold. Among the eight participating Cache schools, only two fall below 60% proficiency. Cache, therefore, represents a distinct implementation context in which mentors serve students identified as needing additional math support in schools where a larger share of students is proficient than in other MM LEAs.

Participating schools also differ markedly in the populations they serve. Figure 5 shows the relationship between the percentage of students who are economically disadvantaged and the percentage proficient in mathematics across participating schools. The association between the percent of economically disadvantaged students and the percent proficient in math on state standardized assessments (RISE) is both strong and negative. Schools in San Juan serve student populations that are almost entirely economically disadvantaged, while several Cache schools serve populations in which fewer than one in four students is economically disadvantaged.

Figure 5. The Relationship Between the Concentration of Students Identified as Economically Disadvantaged and Math Proficiency Rates Across Participating Schools



Sources. STEM AC MM program data, USBE School Report Cards and Data Gateway

Table 1Table 3 provides additional detail on the characteristics of participating LEAs.

Table 3. Characteristics of Schools Participating in STEM AC MM by LEA

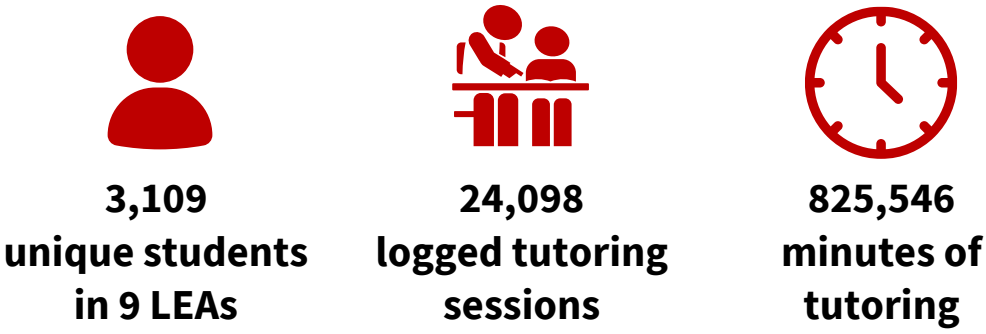
LEA	# Schools	Grades	Total Student Enrollment	Economically Disadvantaged (%)	English Language Learners (%)	Students with Disabilities (%)	Proficient in Math (%)
Cache	8	K-6	4,343	29.4	4.4	17.3	62.2
Millard	2	5-8	961	44.5	6.5	18.5	50.9
FPA	1	K-12	1,241	44.0	19.0	16.0	40.2
Granite	1	K-5	270	66.0	34.0	30.0	32.8
Weber	2	K-9	993	48.5	10.5	17.5	32.1
CCID	1	K-8	329	33.0	3.0	29.0	30.2
GPA	1	K-8	735	55.0	4.0	23.0	30.1
San Juan	3	K-6	550	99.6	23.3	18.7	17.0
UICS	1	6-12	245	96.7	86.0	6.0	–

Note. Percentages reflect the average across participating MM schools in a given LEA if there is more than one school. UICS percent proficiency in 2025-26 was suppressed.
Sources. STEM AC MM program data, USBE School Report Cards and Data Gateway

Students Reached

Participation logs indicate that Math Mentors met and substantially exceeded its Year 4 goal of reaching at least 2,000 students. Mentors provided tutoring to 3,109 students across the 9 participating LEAs, delivering 24,098 documented tutoring sessions totaling 825,546 minutes of tutoring time (13,759 hours; see Figure 6).

Figure 6. STEM AC MM Program Reach in 2025-26



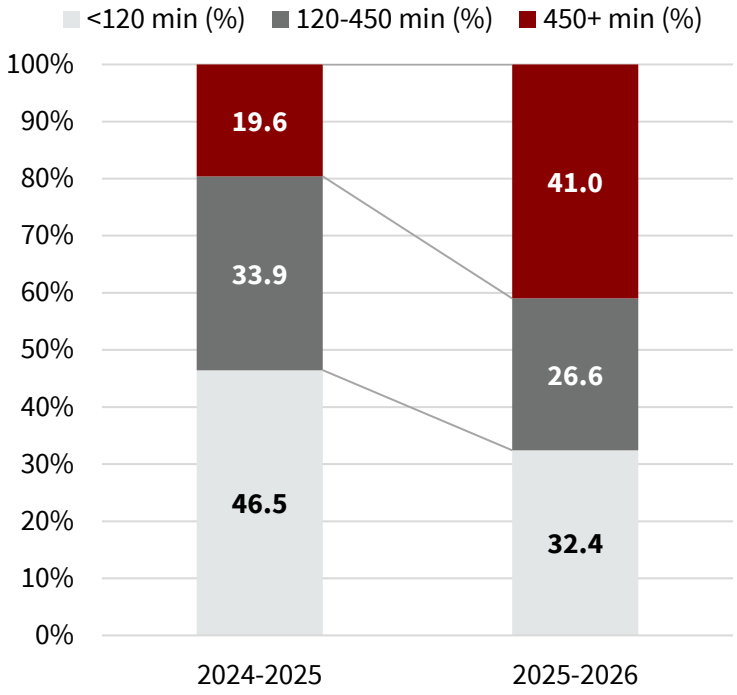
Source: UEPC participation logs

Tutoring Dosage and Intensity

Dosage is the implementation feature most consistently linked to tutoring effectiveness, and it was the primary area for improvement identified in the Year 3 evaluation. While definitions vary, NSSA (2025) recommends that students receive tutoring in three or more sessions per week, for 30 to 60 minutes per session, over at least 10 weeks. This translates to roughly 900 total minutes, or 15 hours, of tutoring over the course of a program year.

The STEM AC MM program made substantial progress toward the goal of increasing tutoring dosage. As shown in Figure 7, the distribution of tutoring dosage shifted markedly between Year 3 and Year 4. The percentage of students receiving at least 450 minutes of tutoring rose from 19.6% to 41.0%. This 21-percentage-point gain was driven almost entirely by the lowest dosage category, as the share of students receiving fewer than 120 minutes fell by 14 percentage points, from 46.5% to 32.4%. Median dosage increased from 125 to 295 minutes, while mean dosage increased from 274 to 726 minutes.

Figure 7. Distribution of Tutoring Dosage in Year 3 (2024-2025) as Compared to Year 4 (2025-2026)



Note. N = 2,127, 2024-2025. N = 3,109, 2025-2026
Sources: 2024-25 and 2025-26 Math Mentors participation logs.

Tutoring dosage, on average, remains below the level at which the strongest effects would be expected (i.e., 900 minutes), and it varies widely across students and sites. Across all participants, dosage ranged from 15 minutes to 9,405 minutes. Table 4 summarizes dosage by LEA.

Table 4. Tutoring Dosage by LEA

LEA	N	%	Mean Minutes	Median Minutes	SD	% (<120)	% (120-450)	% (450+)
CCID	235	7.6	468.6	325.0	560.8	30.2	37.0	32.8
Cache	1,021	32.8	1,089.6	570.0	1,473.3	28.1	17.5	54.4
FPA	249	8.0	721.8	240.0	1,176.7	35.7	29.7	34.5
GPA	312	10.0	581.6	352.5	680.0	28.9	28.5	42.6
Granite	113	3.6	1,391.8	1,380.0	870.4	3.5	13.3	83.2
Millard	293	9.4	483.2	225.0	634.3	33.5	33.5	33.1
San Juan	267	8.6	483.8	170.0	940.2	40.5	36.7	22.9
UICS	129	4.2	827.6	520.0	920.4	20.2	27.1	52.7
Weber	490	15.8	282.9	125.0	362.6	48.0	31.0	21.0
Total	**3,109**	**100**	**726.0**	**295**	**1094.4**	**32.4**	**26.6**	**41.0**

Source: MM program participation logs

Median dosage varied considerably across LEAs, ranging from 125 minutes in Weber to 1,380 minutes in Granite. Granite in particular stands out, with 83.2% of its 113 participating students receiving at

least 450 minutes, and a median dosage exceeding the NSSA benchmark of 900 minutes. Cache, which serves roughly one-third of all program participants, reached 450 minutes for 54.4% of students, suggesting that substantial dosage may be achievable at a larger scale. Weber and San Juan served the largest shares of students at minimal dosage (e.g., < 120 minutes), at 48.0% and 40.5%, respectively.

Further analysis reveals that mentor caseloads varied considerably. Across the 75 mentors, the median mentor documented 288 sessions with 54 students, though caseloads ranged from 5 to 224 students and session counts from 9 to 917 over the course of the program year. These differences identify mentor caseload and service structure as possible factors for further examination. For example, mentors serving very large numbers of students may face logistical barriers that could constrain their ability to meet NSSA benchmarks for recommended total minutes of tutoring and/or small group sizes. However, the descriptive data do not establish whether mentor caseload was associated with student tutoring dosage, nor whether the number of students served across the year corresponded to the size of the individual tutoring groups.

Indeed, the issue of small group size was a commonly cited feature of what mentors discussed as a key success when they had it and, conversely, was among the most commonly cited needs when it was absent. Mentors shared that "*Working with the students in smaller groups, gave me the opportunity to "*tailor*" the tutoring to what they needed the most, in order to become more confident in math*," while school staff also underscored the value of the mentors' work in small groups as, "*providing the one on one help with struggling students that teachers don't have time for*." Other comments highlighted the potential of more mentors to better meet the need, sharing that "*More mentors! Our elementary had two math mentors. we could have used 10! There are so many kids that would benefit from small group math mentoring*." Mentors also named the benefit of being able to reduce group sizes with the addition of another mentor midyear to spread caseload. As one mentor shared:

> *I believe the addition of a math mentor to our school significantly benefited the students in our groups, but specifically because the school chose to split our existing groups into smaller sizes rather than add new students to the newly hired math mentor's schedule.*

Mentors also identified concrete barriers to accumulating dosage, several of which are beyond their control. In focus groups, mentors described school schedules that limited when students could be pulled for tutoring, particularly in smaller schools where elective offerings are more limited. As one mentor explained:

> *Down here... I can't pull students for a group... Then I would be pulling them from music, from library, from art. We don't have a lot of electives down here. So mainly the only time I can see a student one-on-one is during their actual math lesson or during their recess or their lunch, which then they're missing out on the social aspect of school.*

Others highlighted needing "*More time with some students*," while another elaborated that they wished they "*…could pull out the students that need the most help as needed. I don't think I get to work with certain enough during the week.*" Highlighting that mentors are noticing that some students who may need greater intensity to experience the benefits of tutoring. This attention to the intensity and quality of time in tutoring sessions also emerged in mentors' comments about the value of individual

tutoring sessions noting that, "*More implementation of one on one sessions would be amazing for the students*" for individualized support, motivation, and encouragement. As one mentor explained in an open-ended survey response:

> *It seemed like when many students were around their peers, they had a lot of trouble focusing and finding motivation to do their math work - computers can be so distracting when peers are present. Any chance we have to get individual time with students is exponentially more helpful.*

Small-Group Tutoring Was Consistent, but MLS Use and Teacher Communication Were Uneven

Beyond dosage, the STEM AC MM program model specifies that tutoring be delivered individually or in small groups, paired with MLS, and reinforced through communication with school staff and consistent messaging about a growth mindset. Participation logs and end-of-year surveys offer evidence on each.

- **Small-group and individual instruction.** Participation logs indicate that 93% of the 24,098 documented sessions were delivered individually or in small groups, an increase from approximately 85% in Year 3 and a strong signal of alignment with NSSA guidance. More than 4 out of every 5 mentors documented that 85% of their sessions or more were in small groups. Mentor's end-of-year survey reports were consistent: 98% said they met with students individually or in small groups every week.
- **Integration of math learning software.** Software use was documented in 52.9% of tutoring sessions, up from fewer than half in Year 3, but still the least consistently implemented element of the program model. Variation across LEAs and mentors was substantial, ranging from 0 to 100%. While 21 mentors documented software use in at least 80% of their sessions, another 14 documented it in fewer than 25%. Mentor's survey responses showed a similar pattern: one in four reported that they interacted with students as they used MLS never or only some weeks.
- **Growth mindset messaging.** Mentors reported encouraging a growth mindset with a high degree of consistency (93% every week).
- **Communication with teachers and school staff about student performance.** It was the second least consistent practice among mentors, with 58% reporting weekly communication. Mentors often discussed this as an area for improvement, noting that this communication could be improved by "*more direct communication to teachers would be helpful, rather than having word passed down the line through site supervisors or mass staff email communication*." Others suggested it would be beneficial for "*…the teachers to receive a little more direction of how the mentor could be used in their classroom*," to facilitate more effective use of mentor time. Mentors in focus groups also described active coordination with classroom teachers as central to effectively targeting their support (additional details in Appendix A), suggesting that the variation reflects differences in school structures as opposed to mentor priorities.

Mentors Rated Training and Support Positively and Identified Areas to Improve

This section covers how the program prepared and supported mentors in Year 4. It begins with the two trainings offered to new mentors, including how mentors rated them, what they valued most, and what they wanted improved. It then turns to the ongoing support mentors received, from STEM AC resources and site supervisors to the participation log, and how mentors rated that support across schools. It ends with mentors' overall experience in the program.

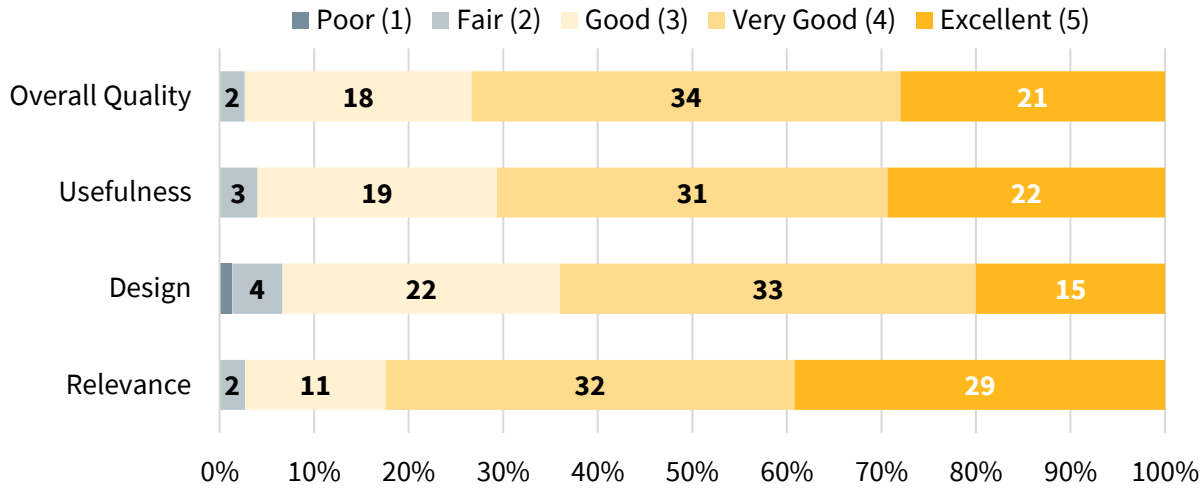
Mentor Training

In addition to monthly newsletters, training, and updates provided to all mentors in Google Classroom, new mentors were invited to complete two training sessions: (1) in-person training provided by a STEM AC staff member, and (2) Saga Coach training, an online, self-paced professional learning program developed by Saga Education that prepares mentors to build strong relationships, promote student engagement, set clear expectations, and support academic growth. Saga Coach training was optional for returning mentors.

Among the 53 unique mentors who responded to either or both of UEPC Mentor Training Exit Surveys (*N* = 75 total), 22 completed the survey following both trainings, 20 following only the training provided by STEM AC staff, and 11 following only Saga Coach training. Because ratings did not differ significantly between training types, analyses of Likert-scale items combine results across both.

Mentors' Overall Impressions of Training. First, respondents to the UEPC Mentor Training Survey were asked to provide ratings for four items assessing their perceptions of the quality, usefulness, design, and relevance of the training. Respondents rated all four items on scales that ranged from 1 ("poor") to 5 ("excellent"). As seen in Figure 8, mentors reported positive experiences with the quality of the trainings. Mean ratings ranged from 3.76 to 4.19 on the five-point scale. These means are all above the scale midpoint, between "good" and "very good."

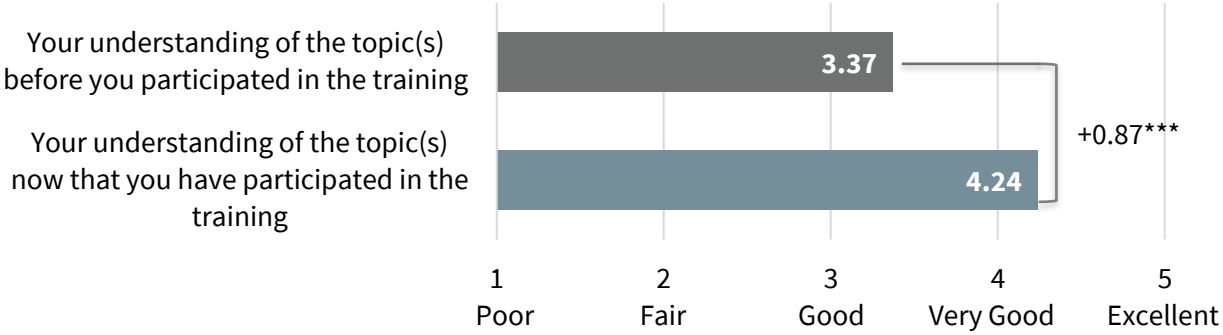
Figure 8. Mentors' End-of-Session Ratings of the Trainings



Note. *N* = 75. Values display counts of participants.
Source: UEPC Mentor Training Exit Survey

Perceptions of Changes in Topic Understanding. Second, survey respondents rated two items assessing their understanding of the objectives and covered topics before and after participating in the training. Respondents rated the items on a scale from 1 ("poor") to 5 ("excellent"). Mean ratings are provided in Figure 9. As shown, ratings reflected statistically significant gains in understanding from before training (mean rating = 3.37) to after training (mean rating = 4.24).

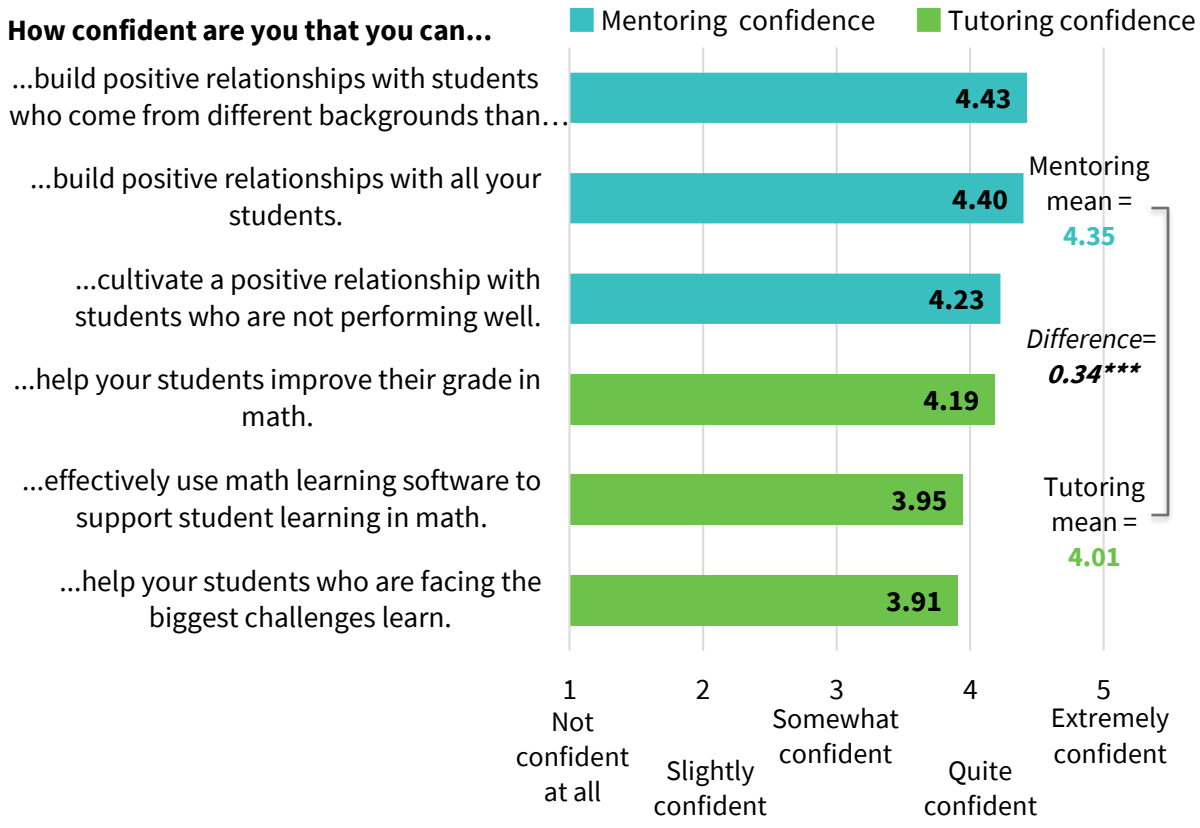
Figure 9. Mean Ratings of Mentors' Understanding of Topics Before and After Training



*p < .05. ** p < .01. *** p < .001.
Note. N = 75.
Source: UEPC Mentor Training Exit Survey

Perceptions of Confidence in Mentoring and Tutoring Competencies. Third, survey respondents were asked to rate their confidence in key tutoring and mentoring competencies, including building strong relationships with students and supporting academic growth. Mentors rated each skill on a scale from 1 ("not confident") to 5 ("extremely confident"). Figure 10 shows the mean score for each item along with the group mean for the items addressing mentoring and tutoring.

Figure 10. Perceptions of Mentoring and Tutoring Competencies



Note. $N = 75$. *$p < .05$. ** $p < .01$. *** $p < .001$.
Source: UEPC Mentor Training Exit Survey

As seen in the figure, mentors expressed high levels of confidence across the mentoring items (group mean = 4.35), but results were more varied for confidence in mentors' math tutoring abilities. Although mentors were confident in their ability to help students improve their grades in math, they were notably less confident in their ability to use MLS to support students' learning and help students facing the greatest challenges. Indeed, the group mean for mentoring competencies was significantly higher (+0.34 Likert scale points) than the group means for tutoring competencies. Confidence was lowest for using MLS to support students' learning and for helping students facing the greatest challenges, the two competencies most directly tied to the academic side of the mentor role.

Most Valuable Aspects of Mentor Training. Fourth, survey respondents were asked to answer an open-ended item assessing their perceptions of the most valuable aspects of their training. Across both trainings (i.e., STEM AC staff-led and/or Saga Coach training), mentors most frequently identified growth mindset strategies and relationship-building with students as the most meaningful and valuable aspects of their training experience. Rather than simply noting these were topics covered in the training, mentors emphasized that these strategies directly shaped how they planned to interact with students, supported persistence, and fostered confidence. Mentors indicated that they valued learning how to use specific growth mindset language to reinforce effort, normalize mistakes, and help students view challenges as opportunities for learning. They also appreciated learning about

how to build trust and rapport with students, noting that examples and videos made these strategies feel real and actionable.

However, some differences in what participants valued and what they felt needed to be improved emerged across both the STEM AC provided training and the Saga Coach training.

Most Valuable Aspects of Training Provided by STEM AC Staff. Respondents frequently emphasized that growth mindset content was impactful, shifting how they think about student learning and their own role as a mentor.

> *I found the growth mindset information the most interesting and valuable. Learning about the difference between growth mindset and fixed mindset helped me to understand that how we think about learning new things and the challenges that brings really affects how we respond.*
>
> *The portions on growth mindset and neuroplasticity. I already expected the portions on assisting the students with math, keeping records, positive reinforcement and praising effort over achievement etc. but just loved the emphasis on helping students develop that growth mindset and have a supporter, a "mentor" helping them in a way that can contribute throughout their life.*
>
> *I found the mindset training very helpful and strategies to help the students change their mindset. I also appreciated learning how to fill out the mentor log.*

Mentors also appreciated tools that clarified expectations and helped them feel confident in getting started, such as guidance on mentor responsibilities and how to track student progress. For example, mentors shared in open-ended responses about what they found most valuable about the STEM AC training included how it supported "*Clarifying and specifying our goals as math mentors,*" and left feeling like "*my job expectations were clear, I know what my purpose within the school is*." Others called out the value of the training on using the participation log and the access to videos and resources.

Most Valuable Aspects of Training Provided by Saga Coach. Respondents frequently emphasized the value of learning about strategies to build positive relationships with students and maintain engagement.

> *I really enjoyed the relationship part of the training and how building a good, supportive relationship helps students feel more confident in learning something new and challenging.*
>
> *The fact that the relationship I have with the student can be the most important factor in tutoring. The reminder that the ratio of work should be 30 me, 70 student, and to maintain high expectations of all students!*
>
> *This training provided some support in maintaining a ratio of fun to learning. It introduced some fun new ideas to keep my students engaged.*

Respondents highlighted questioning techniques, interaction pacing, and practical examples that helped them envision how to connect with students in a supportive and motivating way. For example:

> *I think that the reminders to give wait time, ask rigorous questions, and to keep a "yet" mindset were very helpful.*

I loved some resources it gave, such as what to do in certain situations, things to say, and so on. Those were really valuable and I'll definitely be using them.

The training gave a lot of good tips and ways to improve sessions and make them high quality each time. I really liked the examples it gave of rigorous questions and ways that we can push our students with high expectations.

Areas Identified for Training Improvement. Finally, survey respondents were asked to respond to an open-ended survey item about their perceptions of how the training might be improved. Mentors' responses again differed by training type, reflecting differences in how the content was experienced and where additional support was needed.

Areas Identified for Improvement in Training Provided by STEM AC Staff. Requests for improvement primarily focused on logistics and implementation support, not the content of the training. Mentors asked for opportunities to observe experienced mentors, additional guidance on how to structure sessions, and clearer direction on how to collaborate with teachers or identify which students to support.

I haven't started as a math mentor yet, but I would love to watch a few sessions with someone that has done this before. Doing a shadow session with a mentor, or a teacher do a session with a student would be helpful.

I think maybe helping us know how to structure a session would be great. Some examples of how to make our 20-minute sessions the most effective in terms of scheduling would be helpful.

I would like the training to involve a little more how to work with teachers in relation to the math mentor role all teachers have their own visions for their students and while we take them out and work with them ultimately it's the teacher who says what they need and what they want from us with many differing opinions.

Areas for Improvement in Training Provided by Saga Coach. Here, requests for improvement focused more on pacing, platform usability, and tailoring content to diverse student populations. Mentors specifically asked for different pacing to avoid feeling overwhelmed, clarification of seemingly contradictory guidance about how and how often to engage students in conversations about their lives outside of math, and concrete strategies for working with math software and with students who are disengaged or reluctant to participate. The following representative quotes highlight areas for improvement mentors identified in the Saga Coach training:

I would like to slow down a bit and practice some of the techniques I have learned and then learn more. I'm a little overwhelmed with all the information at once.

I was confused about how to structure our tutoring sessions. In some cases, it told us to get to know students, but in other cases, it told us to not talk too much about things that don't have to do with math.

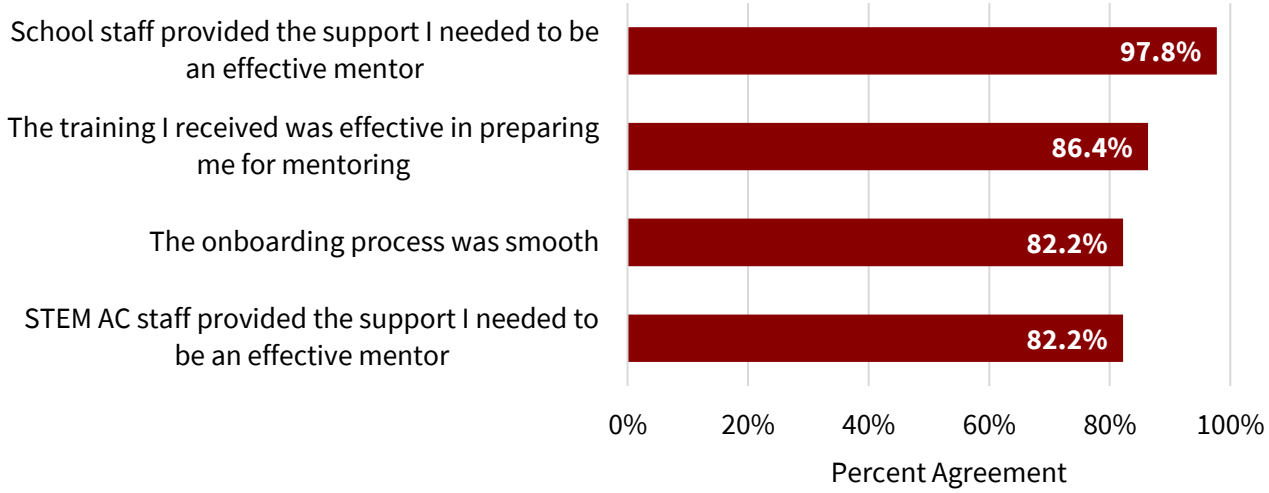
The training seems geared to older students and students who are trying. I work with grade school kids, and they DON'T want to be there. They are embarrassed, inattentive, and just not motivated. They're not worried about consequences.

I think it's great as is. Maybe it could include more about working with students with learning behavior problems and how to navigate being able to properly teach them, or different methods to try.

Ongoing Support for and Monitoring of Mentoring

Mentors in Year 4 were supported through onboarding and training provided by STEM AC, monthly resources and newsletters distributed through Google Classroom, a newly established lead mentor role, day-to-day supervision by site supervisors, and monitoring through the UEPC-managed participation log. Mentors rated the supports available to them positively overall, though with a clear ordering. As shown in Figure 11, mentors were most likely to agree that school staff provided the support they needed to be effective (mean = 4.64, 98% agreement). Ratings of onboarding (4.18), STEM Action Center support (4.18), and training effectiveness (4.14) were also positive but consistently lower.

Figure 11. Mentors' Perceptions of Support



Note. *N* = 45. Percentages reflect the percent of respondents who agreed (4) or strongly agreed (5) with the item.
Source: Math Mentors EOY Survey.

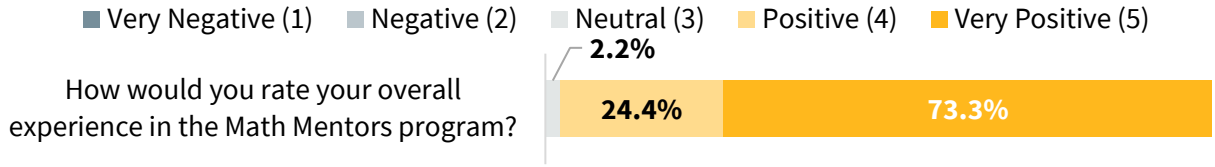
Mentors were hired by districts, placed in schools, and embedded in school teams, so their experience of support depends on the local context that varies across the 20 participating schools. Findings from the mentor focus groups speak to this variation. Mentors described schools where teachers and administrators were unclear about the mentor role, sometimes treating mentors as general classroom aides, alongside schools where sustained administrative attention had produced shared understanding over several years. As one mentor put it, "*They weren't sure what the difference between a math mentor was and an instructional assistant...*"

Mentors' own suggestions centered on structures that would improve upon this lack of shared understanding, including designating lead mentors at each school, creating opportunities for mentors to learn from peers across schools, and clarifying mentor roles and expectations with school teams.

Despite these variations, mentors' overall assessment of the program was strongly positive. Among mentors completing the end-of-year survey, 97.8% described their overall experience with MM as positive or very positive (see

Figure 12).

Figure 12. Mentor Overall Experience in the MM Program



Note. *N* = 45.
Source: Math Mentor EOY Survey

Conclusions and Limitations for Implementation Findings

Overall, the fourth year of the MM program demonstrated meaningful growth in both program reach and progress toward recommended program design goals (e.g., small group, higher dosage). The UEPC analyses found that the program exceeded its staffing and student targets, placing 75 mentors in 20 schools across nine LEAs and delivering 24,098 documented tutoring sessions totaling 13,759 hours to 3,109 students. The share of students receiving at least 450 minutes of tutoring rose 21 percentage points over Year 3, comprised almost entirely of students who would previously have received less tutoring. Mentors delivered 93% of sessions individually or in small groups, and the program continued to serve schools with substantial academic need, with math proficiency below 60% in eight of the nine participating LEAs.

Implementation, however, remains varied across LEAs, schools, and mentors. Median tutoring dosage ranged by an order of magnitude across LEAs, and most students still received less tutoring than the level associated with the strongest effects. MLS was incorporated into tutoring sessions as part of the stacked intervention approach in roughly half of logged tutoring sessions and varied across mentors and sites. Mentors reported strong preparation for the mentoring and relationship-building aspects of their role, but less confidence in the tutoring and academic aspects, particularly in supporting students facing the greatest challenges and in effectively using MLS to support learning. Mentors also experienced support unevenly, rating school-based support considerably higher than program-level training and support.

There are several limitations to consider in interpreting these findings. First, participation logs document tutoring that mentors recorded and do not capture the full range of mentors' activities at MM schools, and anecdotal evidence suggests that tutors may not be logging all sessions, so dosage figures likely represent a lower bound on tutoring delivered. Second, survey findings are self-reported and subject to bias, and item-level averages can obscure variation. Although data collection participation improved, survey coverage remains uneven across sites, most notably for the school staff survey, where responses were concentrated in a small number of LEAs. Focus groups reached mentors from seven of the nine participating LEAs, so the perspectives of mentors in Millard and San Juan are not represented. Finally, Year 4 was the first year the UEPC designed and managed the

participation log, so year-over-year comparisons of documented tutoring reflect some combination of additional tutoring delivered and an improved approach to documentation.

Outcome Evaluation Findings: Change in Student Attitudes

In this section, we examine whether participation in STEM AC MM was associated with positive changes in students' attitudes toward mathematics, addressing the third evaluation question.

> **EQ3.** Is participation in MM associated with positive changes in students' mathematics attitudes?

Strengthening students' confidence in mathematics is a central aim of the STEM AC Math Mentors program. Attitudes toward mathematics are consistently linked to engagement, persistence, and achievement (e.g., Gottfried et al., 2007; Korpershoek et al., 2020; McLeod, 1994). To answer this evaluation question, we drew on multiple sources of evidence across respondent groups. Students provided self-reports of their attitudes in the UEPC EOY Student Survey; mentors reported on the changes they observed in their students through closed and open-ended items on the UEPC EOY Mentor Survey, through focus groups, and through voluntarily shared success stories submitted to STEM AC MM program leaders; and participating MM school staff reported their observations on the UEPC EOY Staff Survey. Throughout this section, we present student findings first and then examine whether mentors and school staff independently describe similar or different patterns. Students are the direct source of information about their own attitudes, so mentor and staff reports serve as a check on what students self-reported.

Student Survey Respondents

Student survey respondents are distributed unevenly across LEAs and more evenly across grade levels. These patterns shaped our approach to analyses, allowing for comparisons by grade but not LEA. The heat map in Figure 13 visualizes this distribution of respondents by LEA and grade level.

Figure 13. Student Survey Respondents and Response Rates by LEA and Grade

	Grade/Grade span							
LEA	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th-12th**	**Percent**
Cache	52	73	67	72	0	0	0	37.7%
Weber	23	15	21	15	91	27	53	35.0%
GPA	24	11	2	4	14	2	0	8.1%
Millard	0	0	19	19	14	0	0	7.4%
San Juan	13	15	0	0	0	0	0	4.0%
UICS	0	0	1	7	10	10	0	4.0%
CCID	2	3	0	2	2	3	0	1.7%
Granite	1	3	0	0	0	0	0	0.6%
FPA	0	0	0	0	1	0	0	0.1%
Percent	16.4%	17.3%	15.9%	17.0%	19.0%	6.0%	7.7%	

Note. $N = 700$. Cells contain counts of student responses by each LEA and grade. Percentages do not sum to 100 due to a small number of student responses missing either LEA or grade.
Source: MM Student EOY Survey

Coverage was reasonably balanced across grade levels from 3rd to 7th grade, with each grade from third through twelfth accounting for between 6% and 19% of responses. Coverage across sites was more uneven. LEA response rates ranged from 0.4% to 50.0%, with Weber and UICS overrepresented relative to their share of program participants and Cache and FPA underrepresented. The analyses below primarily use unadjusted descriptive statistics with selected analyses that examine differences by students' grade level while controlling for LEA rather than reporting results by LEA.

Measures

Each of the four mathematics attitudes (i.e., confidence, effort, interest, valuing, and identity, and growth mindset) was assessed with three items rated on a five-point scale and averaged to form a composite score at each time point (i.e., retrospective pre- and post-items). An additional set of two constructs was measured as post-only scales and assessed students' satisfaction with working with their mentor and satisfaction with the MLS, if they used it with their mentor. We assessed internal consistency using reliability statistics and found that all scales exceeded the thresholds, indicating that the items measured the same underlying factor (e.g., Cronbach's alpha > 0.70; see Appendix C for the full reliability statistics).

Confidence. Three items assessed students' confidence in their mathematical ability, such as how confident they were that they could learn all the material presented in math.

Effort. Three items assessed how much effort students put into mathematics, such as how hard they tried to learn the math they were taught.

Interest, valuing, and identity (IVI). Three items assessed students' interest in mathematics, the importance they placed on doing well, and the degree to which they saw themselves as capable in mathematics.

Growth mindset. Three items assessed students' belief that mathematical ability improves with effort, such as how much they believed they could learn math even when it felt hard. This reflects an improvement over Year 3, when growth mindset was measured with a single item.

Satisfaction with working with a mentor. A set of three post-only items assessed how much students liked working with their mentor and how much their mentor helped them improve their skills and confidence.

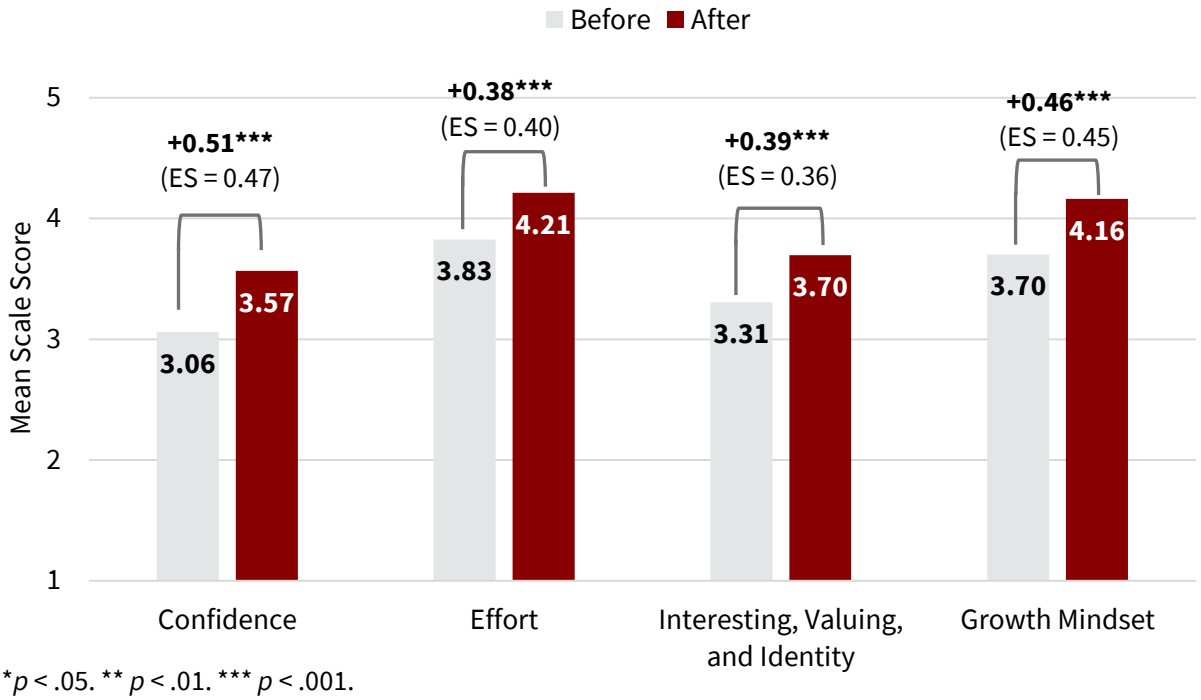
Satisfaction with using MLS. Three parallel post-only items asked the same questions as above about using MLS with their mentor, with an option to indicate that software was not used during mentoring.

Changes in Students' Mathematics Attitudes

As shown in Figure 14, students reported statistically significant positive changes across each of the four scales. Gains ranged from 0.38 to 0.51 points on the five-point scales, with effect sizes between 0.36 and 0.47. Interpreted against appropriate conventional benchmarks for effect sizes (Cohen,

1988), these are small to medium effects.[6] Scale and item-level descriptive statistics are provided in Appendix C.

Figure 14. Pre/Post Change in Students' Mathematics Attitudes



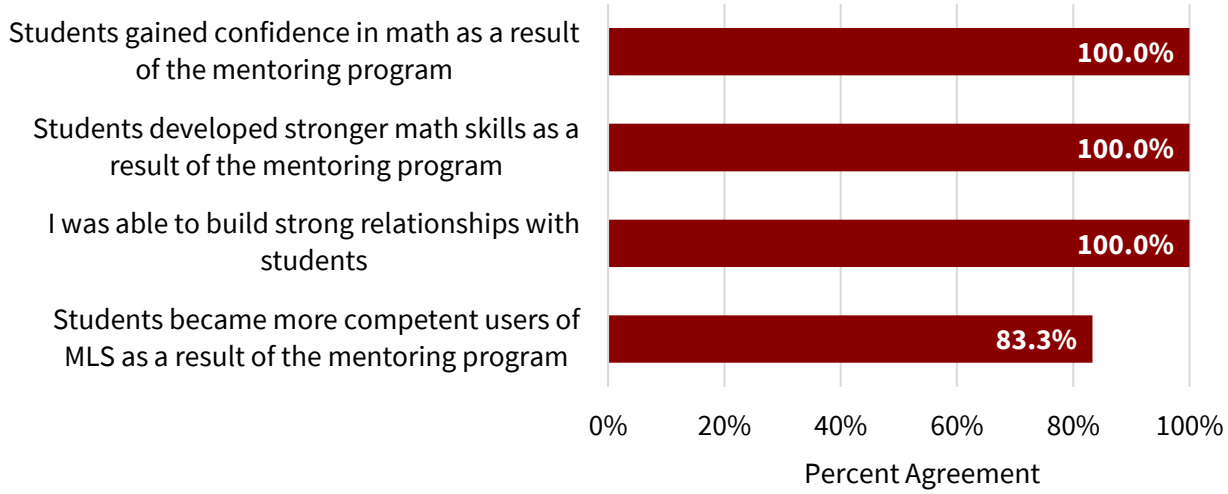
*p < .05. ** p < .01. *** p < .001.
Note. N = 693-698, varies by scale and time point. ES = Effect Size.
Source: UEPC MM Student EOY Survey

Notably, students' confidence in math showed the largest positive change. Students rated their pre-program confidence lowest of the four constructs at 3.06, near the scale midpoint, and reported the largest gain at 0.51 points. This gain was significantly larger than gains in effort (p < .001) and in IVI (p < .001). Growth mindset gains were comparable in magnitude at 0.46 points and also exceeded gains in effort and IVI. Confidence and growth mindset are the two constructs the MM program most directly targets, and both showed the largest gains.

We find that mentors described the same pattern of changes in student outcomes. Every mentor (N = 45; see Figure 15) who completed the end-of-year survey agreed or strongly agreed that the students they worked with gained confidence in math and developed stronger math skills, with mean ratings of 4.71 and 4.69 on a five-point scale.

[6] Effect sizes express the size of a change in standardized units so that results can be compared across measures using different scales. We follow the conventional benchmarks proposed by Cohen (1988), where roughly 0.2 is small, 0.5 is medium, and 0.8 is large. The alternative benchmarks for educational research proposed by Kraft (2020), under which effects above 0.20 are considered large, were developed for causal studies of interventions measured with standardized achievement tests. Because the outcomes here are students' self-reported attitudes, measured close to what mentors work on directly, effects are expected to be larger, and Cohen's benchmarks are the more appropriate benchmark for interpretation.

Figure 15. Mentors' Perceived Impact on Students



Note. *N* = 45. Percentages reflect the percent of respondents who agreed (4) or strongly agreed (5) with the item.

We also asked mentors to respond to an open-ended item about what made the program work well for students. Mentors most often described relationship building and small-group instruction as the conditions that made confidence gains possible. One mentor wrote that working in small groups or one-on-one "*was a great way to help them become more confident in themselves and their learning abilities*," adding that "*having someone believe in them made all the difference for some students.*" Another mentor described the combination directly, writing that small groups allowed them to "*recognize skill gaps and address them one on one*," and that daily conversations helped "*enforce a growth mindset*" while reducing "*the anxiety that is around math*."

Mentors we spoke with in focus groups also described intentionally building up students' confidence with math as a core component of their role, not just as a byproduct of the tutoring. As one mentor explained (see Appendix A for a more full report on the mentor focus groups):

> *I believe that my role as a math mentor is to give confidence and comfort in students as they're doing math. A lot of them are at least a little scared of math or being wrong. And so, I think I'm just there to make sure that they know they can do it themselves, even if they don't believe in it.*

School staff who responded to the end-of-year survey reported the same ordering of impacts, with the highest ratings for relationships and confidence. Because staff responses were concentrated in a small number of sites, we cannot determine whether this pattern holds across LEAs. School staff survey results are provided separately in Appendix D.

Mentor Success Story: "I Almost Like Math Now"

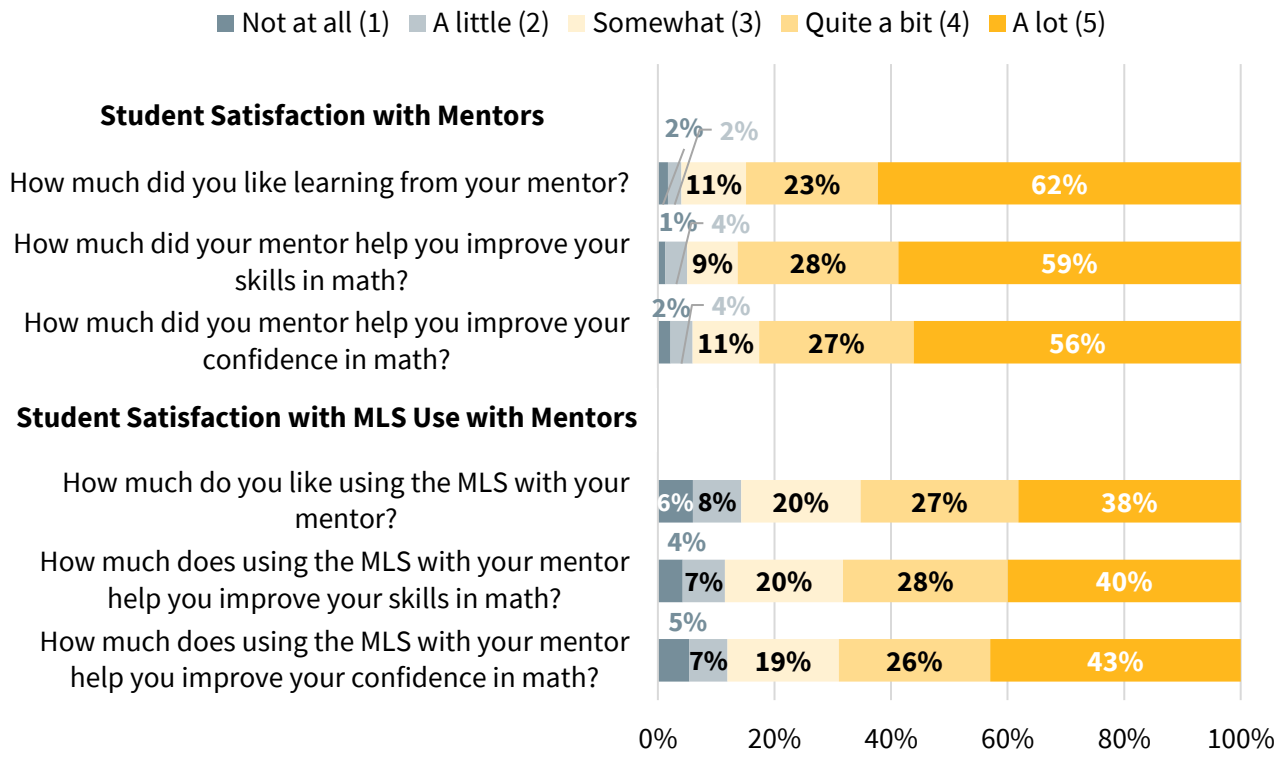
The following vignette shared with STEM AC from a mentor at an elementary school highlights how students' attitudes towards math through the course of tutoring in the MM program:

*I started working with a certain fourth-grade girl on day one of this school year. On that day, when I was explaining my role as a math mentor, she said, "*I just need you to know that I don't like math and you're not going to change that.*" Well, I have worked with her almost daily since that day. At first, I was lucky to get anything out of her. I continued to work with her and help her to see that she could do much more than she thought. I often wondered if I was doing her any good, but I wasn't going to give up. By Christmas, I could see that she was kind of warming up to me and was putting in more effort. About two weeks ago, she caught me off guard. As I was getting ready to go to my next assignment, she stopped me and said, "*You can't tell anyone, but I think I might like math a little bit.*" I asked what was changing her mind about math. She said, "*You make this fraction stuff pretty easy for me.*" I looked at her and said, "*Your secret is safe with me.*"*

Students' Experiences with Mentors and Mentoring and Math Learning Software

Consistent with the broadly positive experiences between mentors and students in MM tutoring reported in the prior section, students also gave high satisfaction ratings of their mentors (Figure 16). Approximately six in seven students reported that they liked learning from their mentor "quite a bit" or "a lot" (84.8%), that their mentor helped them improve their math skills (86.3%), and that their mentor helped them improve their confidence in math (82.6%). Though still generally positive in sentiment, students' ratings of their use of MLS (Math Learning Software) with their tutors were notably lower among those who used MLS during tutoring sessions. Mean ratings were 4.37 for mentoring and 3.92 for MLS, a difference of 0.45 Likert scale points ($p < .001$).

Figure 16. Students' Ratings of Mentoring and Math Learning Software



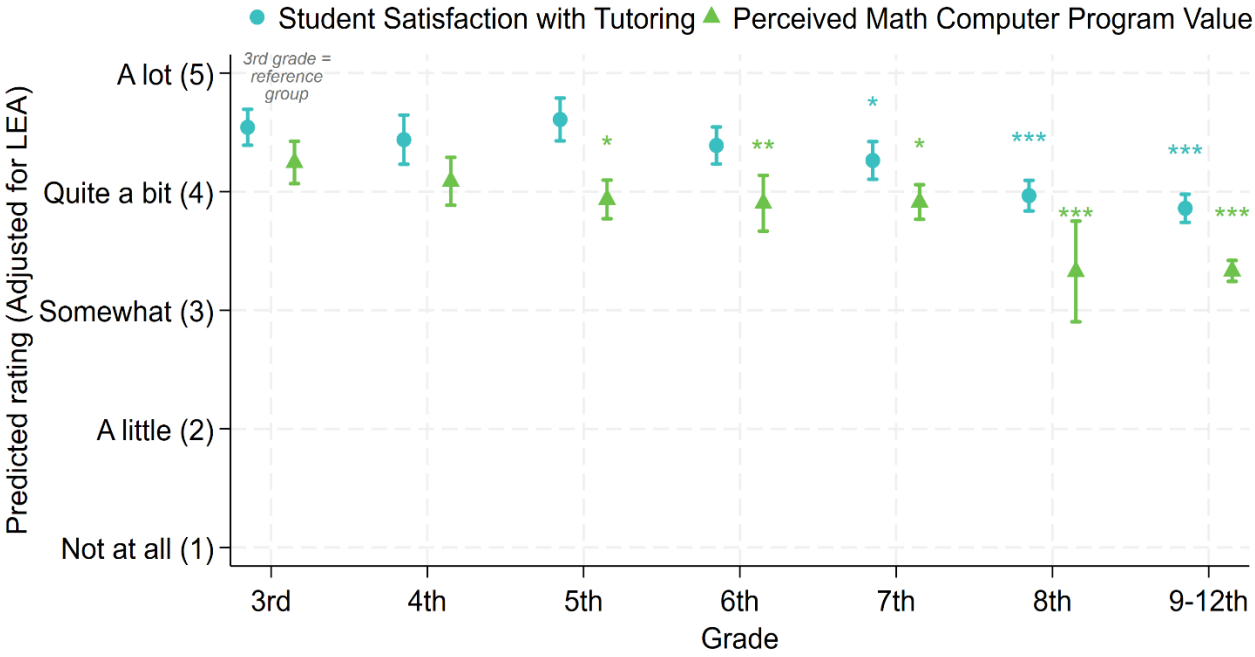
Note. Mentor satisfaction, *N* = 696. Perceived MLS Value, *N* = 552.
Source. Math Mentors EOY Student Survey

Of the 700 respondents, 552 answered the MLS items. While there was an option to indicate that they did not use MLS with their mentor, most students who did not respond skipped the items. We cautiously interpret non-response on these items as indicating they did not use MLS with their mentor, consistent with participation logs showing MLS use in 52.9% of sessions. Mentors likewise rated their own impact on students' MLS competence lower than their impact on students' confidence and skills (Figure 15; 4.33, 83.3% agreement), and the limited evidence from school staff followed the same pattern (Appendix D). Responses from students, mentors, and school staff indicate that MLS use is currently the weakest aspect of the stacked intervention, even as all three groups rate their experiences with the MM program positively overall.

Examining Differences by Grade Level

Responses to the MM Student EOY survey included sufficient participation across grade levels, supporting further examination of student outcomes and program satisfaction by grade level. We used multivariate regression models that controlled for students' LEA (e.g., LEA fixed effects) and clustered standard errors at the school level to assess differences in change scores and satisfaction ratings by grade. Grade was treated as a categorical variable with third grade serving as the reference group. Due to the survey's structure and the smaller number of older student participants, high school students are grouped into a single category. Figure 17 presents predicted ratings of mentoring and software by grade level, adjusted for LEA, with standard errors clustered at the school level. Full model results appear in Appendix C.

Figure 17. Predicted Ratings of Mentoring and Math Learning Software by Student Grade Level



*p < .05. **p < .01. ***p < .001, relative to 3rd grade.
Note. N = 696, Satisfaction with tutoring. N = 552, Computer Program Value. Predicted values from LEA-adjusted regression models; error bars show 95% confidence intervals.
Source: Math Mentors Student Survey

Older students, particularly those in eighth grade and above, rated both mentoring and software less positively than third graders, and the differences widen with grade level. Eighth graders rated mentoring 0.58 points lower and software 0.92 points lower than third graders, each equivalent to a large effect size (respectively, SD = .73; SD = .89). Students in ninth through twelfth grade rated mentoring 0.68 points lower and software 0.92 points lower. However, decreases in ratings of MLS use in mentoring emerge earlier, reaching significance by fifth grade, and are larger than the corresponding differences in mentoring ratings at every grade level.

The magnitude and consistency of these differences, particularly in students' ratings of MLS use in mentoring, suggest they are practically meaningful. At the same time, older students reported larger gains in other areas. Growth mindset gains were significantly larger among eighth graders (+0.38, p < .05) and students in grades nine through twelve (+0.38, p < .01) than among third graders. These older students also began the year with lower growth mindset scores, so the larger gains represent movement toward the levels reported by younger students.

Summary and Limitations for Changes in Student Attitudes

Students who participated in MM tutoring reported gains across all four measured dimensions of mathematics attitudes, with the largest gains in confidence and growth mindset, the two constructs the program most directly targets, and the two on which students often began lowest. Students were broadly positive about their experiences with their MM mentors, with more than eight in ten reporting that their mentor helped them improve both their skills and their confidence "quite a bit" or "a lot". These findings are reinforced by triangulating evidence across data sources. Students, mentors, and school staff independently describe the same three patterns. Confidence and relationships are the

areas where respondent groups agree that the MM program has had the greatest influence, with small-group instruction and consistent contact with a caring adult as the conditions that produce it. The incorporation of MLS into tutoring sessions with mentors is currently the weakest element of the MM program and the one area where all three groups are more moderate in their otherwise positive assessments of experiences in the program.

There are several limitations to consider when interpreting these findings. First, all findings are descriptive. Because only participating students were surveyed, there is no comparison group against which to assess how much change would have occurred over a school year in the absence of the program, and the results cannot be attributed causally to participation in the MM program. Survey coverage was also uneven across sites, meaning we cannot accurately determine how well respondents represent the broader population of participating students. Finally, because responses were anonymous, students' survey responses cannot be linked to their participation records. It is plausible that students' perceptions vary with the amount of tutoring they received, and the present data cannot address that question.

Outcome Evaluation Findings: Student Achievement

In this section, we examine the association between participation in the STEM AC MM program and student mathematics achievement during the 2025–2026 school year. The UEPC's evaluation was guided by the following evaluation question:

> **EQ4.** Is participation in MM associated with positive changes in students' mathematics achievement?

The analyses presented in this section are preliminary and focus on three selected participating Local Education Agencies. We selected these LEAs for the preliminary analyses selected based on data availability and tutoring levels. For these preliminary analyses, we combined MM program participation log data with student demographic and mathematics achievement data provided directly by the three LEAs.[7] In the spring of 2027, UEPC will conduct a subsequent, more comprehensive analysis of the relationship between MM participation and student outcomes across all LEAs. That analysis is expected to provide more complete and consistent student-level information across participating LEAs.

[7] Data Privacy Agreements were established between the UEPC and LEAs participating in the evaluation of the STEM AC Math Mentors program. In addition to adhering to the DPA stipulations, the UEPC maintains University of Utah Institutional Review Board approval for conducting research with human subjects. Only authorized UEPC staff may access the data and data are used only for the purposes for which they were collected.
The Utah Education Policy Center has a Master Data Sharing Agreement with the Utah State Board of Education for use of education data for evaluation and research purposes. The UEPC adheres to terms of the Master Data Sharing Agreement, including terms of use, confidentiality and non-disclosure, data security, monitoring, and applicable state and federal laws, and the University of Utah Institutional Review Board provisions, for the purposes of educational evaluations and research. The UEPC also adheres to FERPA regulations to protect student privacy. Consistent with FERPA and data share agreements, the UEPC does not report data for samples with fewer than 10 students. In addition, the UEPC does not report any personally identifiable information.

Analytic Approach for Student Achievement Outcomes

Student records from each of the three initial LEAs were matched to the MM program participation logs to identify students who participated in tutoring and their total tutoring dosage. Among students appearing in the participation logs, 75.4% (1,087 of 1,441) were successfully matched to the district achievement data. These rates are somewhat below the 80% match rate, which USBE generally considers a desirable benchmark. However, the district data used for matching had already been restricted to students with complete baseline and end-of-year achievement scores; therefore, some unmatched students may reflect missing achievement data rather than record-linkage failures. Students in the district data who could not be matched to the participation logs were classified as non-participants.

Of these students, 1,087 participated in MM during the 2025–2026 school year. This sample differs from those used elsewhere in this report. The participation logs documented 3,109 students served across all nine LEAs, and 700 students completed the EOY Student Survey. The achievement analyses draw on data from three LEAs and include only participants matched to district records and with complete baseline and end-of-year scores. Even so, the 1,087 participants represent about 35% of all students served by MM in 2025–26.

Because the results differed across the three LEAs, analyses were conducted separately by LEA rather than pooling students into a single model. Therefore, the results of these analyses should not be used to make comparisons between LEAs and instead considered as results within each analytic sample.

Measures

Participation. Student participation in the program was measured in three ways:

- **Binary Classification (Yes/No):** A binary indicator of whether the student received any MM program tutoring during the year.
- **Dosage Groups:** Consistent with the previous year's analyses, students were classified into four mutually exclusive dosage groups: (1) non-participants, (2) less than 120 minutes, (3) 120–450 minutes, and (4) more than 450 minutes. These categories were used to examine whether associations with achievement differed according to the amount of tutoring students received.
- **Dosage (Total Minutes):** The total number of minutes of tutoring each student received. Students who did not participate were assigned 0 minutes.

Math Achievement. Mathematics achievement was measured using Acadience Math scores for students in grades K-3 and RISE Math scores for students in grades 4-6. For each student, a baseline score representing achievement prior to the relevant period of MM program participation and an end-of-year outcome score were identified.[8] Because students were assessed using different measures and were enrolled in different grade levels, scores were standardized within each LEA, grade level, and assessment type prior to analysis. Standardized scores, therefore, represent student performance relative to other students in the same LEA taking the same assessment at the same grade level.

[8] For the RISE assessment, the baseline assessment is students' prior year EOY math score. For Acadience Math, the baseline score is the students' beginning-of-year score in the same academic year.

Because scores are standardized at each time point, the average for all students in each LEA and grade is always zero, so a group's average describes its standing relative to other students.

Covariates. The primary adjusted models controlled for baseline mathematics achievement, gender, English Language Learner (ELL) status, and grade level, the student characteristics that were provided consistently across the three LEAs. Other characteristics, such as race/ethnicity and low-income status, were available in some LEA datasets but not consistently enough to be included. Models for LEAs with more than one school also included school fixed effects to account for unobservable differences across its participating schools. The full analysis in spring 2027 will draw on more complete and consistently defined student characteristics.

Data Analysis Strategy

Analyses proceeded in three stages to understand both overall patterns and the potential impact of STEM AC MM program tutoring on student math achievement. These three stages included: descriptive analyses, model-based analyses, and a sensitivity analysis.

Stage 1: Descriptive Analyses.
We first examined participation and achievement patterns without statistical adjustment. We summarized participation rates and median tutoring minutes for each LEA and then compared average standardized baseline and end-of-year mathematics scores by participation status and tutoring dosage. These analyses provide an accessible description of the observed achievement trajectories but do not account for differences in students' initial achievement or other characteristics.

Stage 2: Model-Based Analyses.
We next estimated separate regression models for each LEA to examine whether MM program participation and tutoring dosage were associated with end-of-year mathematics achievement, after accounting for students' baseline achievement and available demographic characteristics. The first set of models compared participants with non-participants. A second set compared students in each tutoring dosage category with non-participants. All models adjusted for baseline mathematics achievement, gender, ELL status, and grade level; models additionally included school fixed effects when LEAs had more than one school to account for differences across schools.

Stage 3: Sensitivity Analyses.
Finally, we conducted several analyses to determine whether the substantive conclusions changed under alternative model specifications. First, tutoring dosage was modeled as a continuous measure in place of dosage categories. Second, we used residualized achievement growth as an alternative outcome. This approach examines whether students performed better or worse at the end of the year than would be expected based on their baseline achievement. Third, propensity score matching (PSM) was used to compare students receiving 450 or more minutes of tutoring with similar non-participants. Together, these analyses provide a check on whether the primary findings depend heavily on a particular definition of participation, dosage, or achievement growth.

Results from Descriptive Analyses

Participation Rates and Median Tutoring Time by LEA

Table 5 summarizes the analytic sample and MM participation across the three LEAs. The sample included 4,877 students across 10 schools, of whom 1,087 (22.29%) participated in the MM program. These 1,087 students represent approximately one-third of all students who participated in MM tutoring across the nine participating LEAs during the 2025–2026 school year.

Participation and tutoring dosage differed across LEAs. LEA 3 accounted for the largest number of students and MM participants, with 794 of 3,998 students (19.86%) participating. Participation rates were higher in LEA 1, where 86 of 222 students (38.74%) participated, and in LEA 2, where 207 of 657 students (31.51%) participated.

Among participating students in the analytic sample, the median tutoring dosage was 625 minutes overall. Median dosage varied considerably across LEAs, ranging from 490 to 1,435 minutes. Thus, although LEA 3 included substantially more MM participants, participating students in LEA 1 received considerably more tutoring at the median.

Table 5. Participation Rates and Median Tutoring Time by LEA

LEA	Grade Span	Total # of Students	% MM Participants	Median Tutoring Time (in minutes)
LEA 1	Elementary	222	38.74	1,435
LEA 2	Elementary	657	31.51	490
LEA 3	Elementary	3,998	19.86	605
Total	**Elementary**	**4,877**	**22.29**	**625**

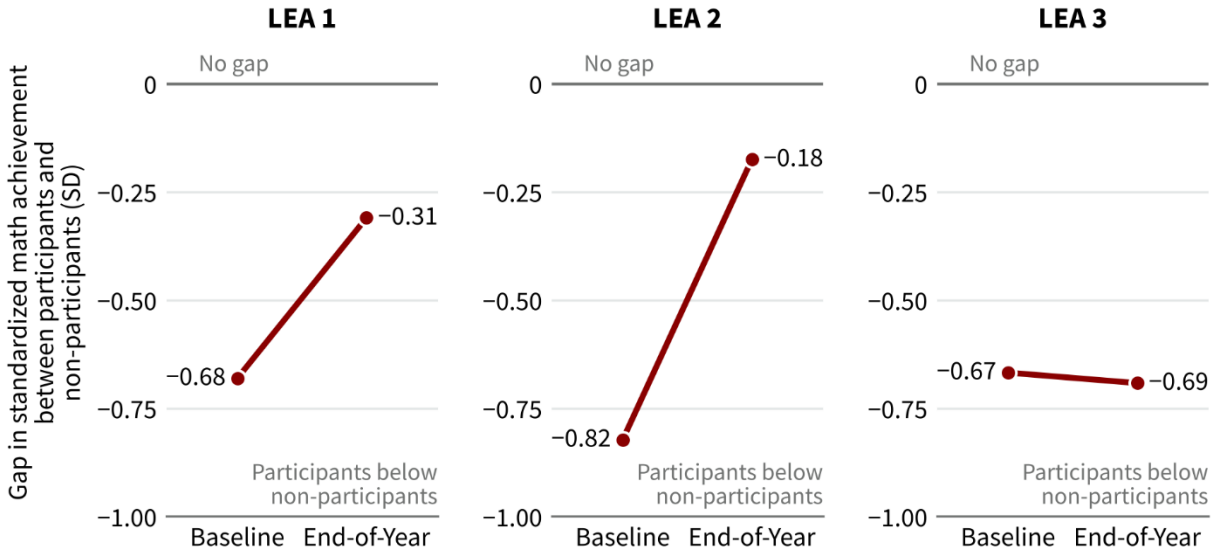
Source: Participation logs

Math Achievement Outcomes by Participation Status

Standardized scores indicate where a student stands relative to other students in the same LEA and grade. They do not show how much a student has learned. Because of how standardized scores are calculated, the average across all students is always zero, both at baseline and again at the end of the year. This means the two groups' averages are tied together, so a change in one group's scores forces a mechanical change in the other group's mean scores. For example, when STEM AC MM participants make gains that close the gap between them and their non-participant peers, the non-participant average moves *down* by a corresponding amount, even if every student's actual test score went *up*. A lower average for non-participants is a result of the calculation and does not mean non-participants learned less or did worse on the test. For this reason, Figure 18 and Figure 19 show only the gap between groups. Because MM serves students who need additional math support, participants start at a lower level than non-participants, so the gap begins below zero. A line that rises toward zero means participants closed part of that gap by the end of the year relative to non-participants; a flat or falling line means the gap stayed the same or widened.

Figure 18 displays the gap in average standardized mathematics scores between MM participants and non-participants at baseline and at the end of the year, separately for each LEA.

Figure 18. Gap In Average Standardized Math Scores Between MM Participants and Non-Participants at Baseline and End of Year, by LEA



Sources: UEPC MM participation logs and LEA student achievement and enrollment data
Note. Values show the difference between participants' and non-participants' mean standardized scores. Values below 0 mean participants scored lower than non-participants on average; values closer to 0 indicate a smaller gap. Scores are standardized within each LEA and grade, so results should be interpreted within each LEA and not compared across LEAs.

As illustrated in Figure 18, a consistent pattern across all three LEAs is that students selected to receive MM tutoring began with lower mathematics achievement scores (i.e., Acadience Math, RISE Math) than students who did not receive MM tutoring. This is expected, given that the program is designed for students who need additional math support. The baseline difference was particularly pronounced in LEA 2. This pattern is important because a simple comparison of participants' and non-participants' end-of-year scores would not provide an appropriate estimate of the association between tutoring and achievement. Participants started with lower scores, so a gap at the end of the year would partly reflect where students began rather than the effect of tutoring.

Math achievement trajectories after baseline, however, differed considerably across the three LEAs. In LEA 3, the gap in mean standardized math achievement between participants and non-participants was essentially unchanged, with participants remaining substantially below non-participants at the end of the year. In contrast, the gap narrowed substantially in LEA 1 and LEA 2, by more than three-quarters in LEA 2 (from 0.82 to 0.18 standard deviations) and by more than half in LEA 1 (from 0.68 to 0.31 standard deviations). Because scores are standardized, a narrowing gap reflects participants gaining ground relative to their peers and does not indicate that non-participants' achievement declined.

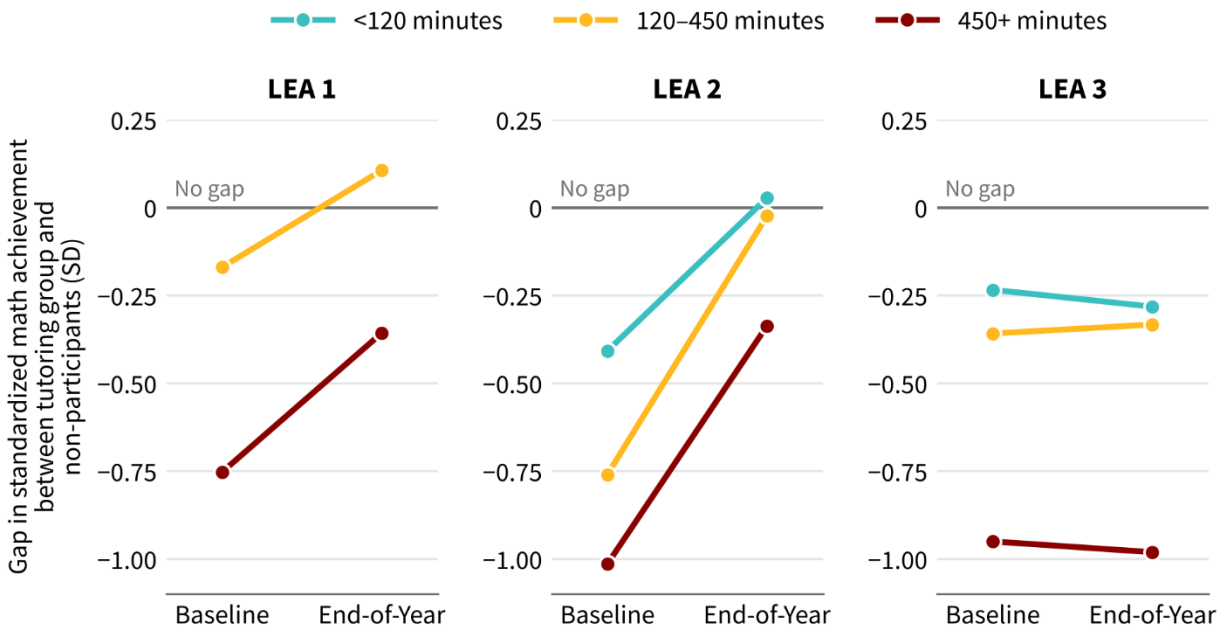
These descriptive results provide initial evidence of different patterns across LEAs: little evidence of improvement associated with participation in LEA 3, but more promising achievement trajectories among MM participants in LEA 1 and LEA 2. Because these comparisons do not adjust for baseline

achievement or student characteristics, the model-based analyses presented in a section that follows provide a more rigorous test of these patterns.

Math Achievement Outcomes by Tutoring Dosage

Figure 19 further disaggregates the descriptive results according to total tutoring dosage. Students are classified as non-participants, participants receiving less than 120 minutes, participants receiving 120–450 minutes, or participants receiving 450 or more minutes of tutoring. Each line shows the gap between a dosage group and non-participants in the same LEA.

Figure 19. Gap in Average Standardized Math Scores Between Each Tutoring Dosage Group and Non-Participants at Baseline and End of Year, by LEA



Sources: UEPC participation logs and LEA math student achievement and enrollment data provided by selected LEAs

Note. Values show the difference between each tutoring dosage group's average standardized score and that of non-participants. Values above 0 mean the group scored higher than non-participants on average; values below 0 mean it scored lower. Scores are standardized within each LEA and grade, so results should be interpreted within each LEA and not compared across LEAs. Results for groups with fewer than 10 students are not reported (LEA 1, <120 minutes). See Table 7 for sample sizes by LEA and dosage group.

As with the participation comparison, the figure demonstrates that tutoring dosage was not randomly distributed across students. In general, students receiving tutoring began with lower mathematics achievement than non-participants, and those receiving the most tutoring frequently had particularly low baseline scores. This reinforces the importance of accounting for baseline achievement when examining the relationship between dosage and end-of-year performance. Because students who received the most tutoring started with the lowest scores, end-of-year differences across dosage groups partly reflect where students began. Comparisons across dosage groups need to account for baseline achievement.

As illustrated in these analyses, the subsequent math achievement trajectories again differ across LEAs. In LEA 3, none of the tutoring dosage groups show a descriptive pattern suggesting that the achievement gap was closed for students who received tutoring relative to non-participants. Students receiving 450 or more minutes had especially low baseline achievement and remained well below non-participants at the end of the year.

A different pattern is evident in LEA 2 and LEA 1. In LEA 2, the gap narrowed substantially for all three tutoring dosage groups, including students receiving 450 or more minutes of tutoring. N LEA 1, the gap in average standardized math achievement closed or narrowed for students in the 120-450 minute and 450 or more minute dosage groups relative to non-participants by the end of the year.

Overall, the descriptive dosage results reinforce the LEA-specific pattern observed in Figure 18. Achievement trajectories were more favorable among tutored students in LEA 1 and LEA 2, particularly at higher levels of tutoring, whereas comparable evidence of improvement was not evident in LEA 3. The adjusted models below examine whether these differences persist after accounting for baseline achievement and student characteristics.

Results from Model-Based Analyses

The descriptive results indicate substantial differences in math baseline achievement between students who did and those who did not receive MM tutoring. We therefore estimated separate regression models for each LEA to examine whether participation and tutoring dosage were associated with end-of-year mathematics achievement, adjusting for baseline achievement and available student characteristics.

Across the analyses, a clear and consistent distinction emerged between the three LEAs. **Results for LEA 1 and LEA 2 provide promising evidence of positive associations between MM tutoring and mathematics achievement, whereas results for LEA 3 do not.** Importantly, this general pattern was robust across multiple definitions of tutoring exposure and of modeling achievement.

Adjusted Association Between MM Participation and Math Achievement

Table 6 presents the adjusted association between MM participation and end-of-year mathematics achievement (i.e., Acadience Math, RISE Math). Estimates represent the adjusted difference in standardized end-of-year math achievement between MM participants and non-participants after controlling for baseline achievement, gender, ELL status, and grade level. The LEA 3 model also adjusts for school.

The results differed substantially across LEAs. In LEA 3, MM participation was associated with lower adjusted end-of-year math achievement (estimate = −0.202, 95% CI [−0.253, −0.151], $p < .001$). Thus, after adjusting for baseline achievement and available student characteristics, the analysis did not provide evidence of a positive association between MM participation and mathematics achievement in LEA 3.

In contrast, participation was positively associated with end-of-year mathematics achievement in both LEA 1 and LEA 2. In LEA 1, participants scored an estimated 0.284 standard deviations higher than comparable non-participants after adjustment (95% CI [0.119, 0.449], $p = .001$). In LEA 2, the adjusted difference was approximately 0.4 standard deviations in favor of MM participants (95% CI

[0.253, 0.547], p < .001). Under Kraft's (2020) benchmarks for education research, effect sizes of 0.20 standard deviations or greater are considered large. Both estimates exceed this threshold.

Overall, the adjusted participation models reinforce the descriptive results: MM program participation was positively associated with mathematics achievement in LEA 1 and LEA 2, but not in LEA 3.

Table 6. Adjusted Association Between MM Participation and End-of-Year Math Achievement, by LEA

LEA	N	MM participants (n)	Estimate	SE	95% CI Lower	95% CI Upper
LEA 1	222	86	**0.284*** **	0.084	0.119	0.449
LEA 2	657	207	**0.400*** **	0.075	0.253	0.547
LEA 3	3,998	794	**-0.202*** **	0.026	-0.253	-0.151

*$p < .05$, **$p < .01$, ***$p < .001$.
Note: Models adjust for baseline mathematics achievement, gender, English Language Learner status, and grade level. Models for LEAs with more than one participating school included school fixed effects. Estimates represent differences in standardized end-of-year mathematics achievement between participants and non-participants.
Sources: Participation logs and LEA student achievement and enrollment data

Adjusted Association Between Tutoring Dosage and Math Achievement

Because participation alone does not capture potentially important differences in the amount of tutoring students received, we next estimated models comparing students across three tutoring-dosage categories with non-participants. Table 7 presents these results. The dosage models again reveal substantially different patterns across LEAs.

Table 7. Adjusted Association Between MM Tutoring Dosage and End-of-Year Math Achievement, by LEA

LEA	Dosage Group (min)	Group n	Estimate	SE	95% CI Lower	95% CI Upper
LEA 1	<120	*n* < 10, not reported				
	120-450	10	0.263	0.195	-0.121	0.647
	450+	74	**0.299*** **	0.089	0.124	0.474
LEA 2	<120	40	**0.327***	0.133	0.065	0.588
	120-450	60	**0.521*** **	0.113	0.299	0.744
	450+	107	**0.359*** **	0.094	0.175	0.542

LEA	Dosage Group (min)	Group n	Estimate	SE	95% CI Lower	95% CI Upper
LEA 3	<120	190	**-0.130****	0.046	-0.221	-0.04
	120-450	149	-0.087	0.052	-0.188	0.014
	450+	455	**-0.274*****	0.033	-0.338	-0.21

**p < .05, **p < .01, ***p < .001.*
Note: Non-participants are the reference group. Models adjust for baseline mathematics achievement, gender, English Language Learner status, and grade level. Models for LEAs with more than one participating school included school fixed effects. Results for groups with fewer than 10 students are not reported.
Sources: Student data provided by selected LEAs and MM program participation logs.

In LEA 1 the estimate for students receiving 120–450 minutes was imprecise and not statistically significant, reflecting the small number of students in this group (n = 10), and too few students received less than 120 minutes to report results, as most LEA 1 participants (86%) received 450 or more minutes of tutoring. Students receiving 450 or more minutes, however, had significantly higher adjusted end-of-year scores than non-participants (estimate = 0.299, 95% CI [0.124, 0.474], p = .001). This finding is particularly notable given the relatively high median tutoring dosage observed in LEA 1.

In LEA 2, we observed statistically significant positive associations across all three tutoring dosage categories. Compared with non-participants, adjusted end-of-year scores were higher among students receiving less than 120 minutes (estimate = 0.327, p = .015), 120–450 minutes (estimate = 0.521, p < .001), and 450 or more minutes (estimate = 0.359, p < .001).

In LEA 3, none of the dosage categories were associated with higher adjusted achievement relative to non-participation. Students receiving less than 120 minutes had lower adjusted end-of-year scores than non-participants (estimate = −0.130, p = .005). There was no significant effect detected for students receiving 120–450 minutes (p = .093). Students receiving 450 or more minutes had lower adjusted scores than non-participants (estimate = −0.274, p < .001). These results provide no evidence that receiving tutoring or greater tutoring dosage was associated with improved mathematics achievement in LEA 3.

Taken together, these models provide further evidence that the relationship between MM tutoring and mathematics achievement varied across LEAs. There was no evidence of a positive dosage-achievement relationship in LEA 3. In contrast, the results for LEA 1 provide evidence of higher achievement among students receiving at least 450 minutes of tutoring, while LEA 2 shows positive adjusted associations across the observed dosage categories.

Sensitivity Analyses

To assess the robustness of these findings, we conducted three additional analyses using alternative definitions of tutoring exposure and student achievement. First, we modeled tutoring dosage as a continuous measure rather than categorizing tutoring exposure (e.g., <120 min, 120-450 min, 450+ min). Second, we estimated models using residualized end-of-year achievement growth as the alternative outcome specification that accounts for students' baseline performance. Third, we used propensity score matching (PSM) to compare students who received at least 450 minutes of tutoring

with non-participant students who had similar characteristics included in the matching procedure. The matching approach was designed to improve comparability between the high-dosage and non-participant groups by accounting for observed baseline characteristics.

The results were highly consistent with the primary analyses. All three analyses produced statistically significant positive estimates for LEA 1 and LEA 2 and negative estimates for LEA 3. Although the magnitude of individual estimates varies by analytic approach, LEA 1 and LEA 2 consistently show evidence of a positive association between MM tutoring and mathematics achievement, whereas LEA 3 consistently does not. Full results appear in Appendix E.

Conclusions and Limitations for Student Achievement Findings

The preliminary 2025–2026 student achievement analyses provide promising, but not uniform, evidence of a positive association between MM tutoring and mathematics achievement. The findings differed substantially across the three LEAs included in this initial analysis.

In LEA 1 and LEA 2, the evidence was consistently favorable. After accounting for students' baseline achievement and available demographic characteristics, MM participants had higher end-of-year mathematics achievement than non-participants. The dosage analyses provided additional evidence of positive associations, particularly for students receiving 450 or more minutes of tutoring in LEA 1. Most importantly, the overall findings for these two LEAs were robust to multiple analytic decisions and remained positive across alternative measures of tutoring exposure and achievement (see Appendix E).

The findings for LEA 3 differed notably. Neither the descriptive nor adjusted analyses provided evidence of a positive association between MM participation and mathematics achievement. Results remained negative or non-positive across participation, dosage, continuous-dosage, residualized-growth, and propensity-score-matching specifications. Analyses conducted separately for each school likewise found no evidence of a positive association. Thus, the difference between LEA 3 and the other two LEAs appears to be a consistent feature of the available data rather than an artifact of a single analytic approach.

These findings should nevertheless be interpreted cautiously. Students were not randomly assigned to receive tutoring, and the descriptive analyses show that students selected for MM generally began with lower mathematics achievement than non-participants. Although the statistical models adjust for baseline achievement and several student characteristics, they cannot account for every potential difference between students who did and did not receive tutoring. The data available directly from the LEAs also limited the characteristics that could be included consistently in the preliminary models. In particular, race/ethnicity and measures of socioeconomic status could not be incorporated consistently because of limitations in the data received from the LEAs. For these reasons, interpret the present findings as evidence of associations rather than definitive causal effects of MM program participation.

Finally, these analyses represent a preliminary examination of 2025–2026 student outcomes. The subsequent full analysis will link MM participation and dosage records to more complete student data

obtained from USBE. Those records will allow for more consistent measurement of student characteristics and will provide an opportunity to determine whether the patterns identified in these preliminary analyses remain when using the more comprehensive statewide student records.

Overall, the preliminary results provide encouraging evidence for MM in two LEAs, where positive associations with mathematics achievement were observed consistently across multiple models and sensitivity analyses. At the same time, the absence of comparable evidence in LEA 3 underscores the importance of examining program outcomes within individual implementation contexts and of using the forthcoming USBE data to further assess the robustness of these findings.

Recommendations

The implementation of the STEM AC MM program showed meaningful progress in Year 4. In particular, tutoring dosage among participants increased substantially, mentor training and support expanded through the lead mentor role and monthly resources, and participation in data collection increased across every instrument. This year's recommendations are grounded in the Year 4 findings and aligned with design principles that support integrating high-dosage tutoring as an intervention to support student learning (e.g., Nickow et al., 2020; Pellegrini et al., 2021; Robinson et al., 2024).

Continue to focus on consistently reaching high-impact tutoring benchmarks for all students

The prior year's evaluation (Y3, 2024-25) recommended ensuring that more students receive tutoring at the intensity associated with stronger outcomes. The MM program made meaningful progress toward this goal, with the share of students receiving at least 450 minutes of tutoring rising from 19.6% to 41%. However, 32.4% of students received fewer than 120 minutes of tutoring this year. NSSA (2025) guidance recommends 900 minutes. Median dosage in Year 4 was 295 minutes, and intensity varied more than tenfold across LEAs, from 125 minutes in Weber SD to 1,380 minutes in Granite SD. This year's newly introduced activity log may reflect both more tutoring and more complete documentation of tutoring efforts.

Mentor caseloads, which ranged from 5 to 224 students over the course of the year, may explain some of the variation we observed in reaching high-dosage tutoring benchmarks. This variation warrants closer attention to understand if it, along with the mentors' service period, group sizes, planned/missed session patterns, contributes to the consistency and amount of time students are tutored. In the broader literature, tutoring effects diminish in large programs as dosage declines and student-tutor ratios rise (Kraft, Schueler, & Falken, 2024; Kraft, Sanderson Edwards, & Cannata, 2024). Additional details about how sites are making decisions on student selection for program participation, mentoring schedules, mentoring assignments, and caseloads would be beneficial to understanding how current program implementation may be facilitated and/or what challenges are faced.

Strategies to support progress in more consistently reaching benchmarks for high dosage tutoring for students across sites include the following.

- Establish a feasible tutoring schedule for each student or group of students, including frequency and session length.
- Using available evidence on caseload, provide guidance and recommendations on mentor caseload size.
- Identify protocols for determining whether the addition of mentors results in more sessions for students participating or more students being served.
- Work with site supervisors at schools with higher and lower median dosages to explore the factors that facilitate or impede reaching high-dosage tutoring benchmarks.

Strengthen the instructional integration of tutoring and use of MLS

The MM program intends to strategically pair high-dosage tutoring with MLS, an approach with promise for addressing the scalability challenges that limit tutoring programs at scale (Bhatt et al., 2025; Guryan & Ludwig, 2023). Although mentors made progress in using MLS in tutoring sessions in Year 4, use remains inconsistent. Software use was documented in 52.9% of tutoring sessions, with mentor-level variation ranging from none to all sessions. Findings from other data sources reinforce this finding and offer insights into why MLS use is limited and uneven. Mentors reported the lowest confidence in using MLS to support student learning among all training competencies, and one in four reported interacting with students during software use never or only some weeks as a part of their tutoring sessions. Mentors and school staff both rated the program's impact on students' MLS competence lower than its impact in any other area. Students rated their experience with the software 0.45 points below their experience working with their mentor, and approximately one in five did not use the software with their mentor at all.

Ratings of MLS use also declined with grade level, with significant differences emerging by fifth grade and reaching nearly one point lower among students in eighth grade and above compared to third graders. Evidence from the STEM AC Digital Math program evaluation (Altermatt et al., 2026) indicates that MLS products vary in the student groups they serve, including by grade span, suggesting that the fit between software and students may contribute to these differences.

To support the quality and fidelity of implementing this stacked intervention approach, the following strategies may strengthen the implementation of math tutoring using MLS.

- Define the appropriate instructional purposes and availability of MLS within a tutoring session, including how to use information about a student's math learning needs to select and support MLS use.
- Determine whether MLS (limited) use reflects confidence in use, availability or access for tutors or students, or other factors.
- Ensure that MM mentors have access to MLS professional learning, training, and technical assistance support (e.g., sample lessons, opportunities to observe MLS integration, coaching) on MLS integration.
- Provide technical assistance to help sites identify MLS programs well suited to the grade levels and math learning needs of the students they serve.

Build a shared understanding of and expectations for the mentor role

Year 4 included changes to how MM program mentors were hired and supported with the move away from AmeriCorps support and the addition of site-based lead mentors. Mentors are hired by districts, placed in schools, and embedded in school teams across participating LEAs. Mentors rated the support they received from school staff more positively than any other source of support (4.64 on a five-point scale), demonstrating that school-level supports are particularly valuable for mentors' work. With this key source of support located within schools, shared understandings of mentors' roles is particularly important for their effectiveness.

Our data reveal that current understandings of mentors' roles and what it means to implement the MM program vary, indicating a need to develop a more coherent, shared understanding of the role across involved stakeholders. In focus groups, mentors described teachers and administrators who were unclear about the mentor role, sometimes treating mentors as general classroom aides, alongside schools where sustained administrative attention had produced shared understanding over several years. Mentors themselves reported communicating with school staff about student performance less consistently than other program features (noting the exception of MLS usage), at 58% weekly. In open-ended responses, mentors suggested program improvements, including better, more direct communication with teachers and clearer guidance for teachers on how mentors can be used. Interruptions to scheduled tutoring included instances in which teachers indicated that students did not need to be pulled on a given day, suggesting that some lost tutoring time reflects differing understandings of the program's purpose rather than scheduling conflicts alone.

These findings echo Year 3, when lead mentors reported that mentors were sometimes assigned tasks outside their intended scope and that some administrators were unfamiliar with the program's design. The persistence of this pattern suggests a need for attention to facilitate change.

Strategies to develop a shared understanding of and expectations for the MM mentors' role across sites include the following.

- Develop a brief, shared description of the mentor's instructional role, including how tutoring supplements classroom math instruction, mentor tutoring responsibilities, features of and evidence for high-impact tutoring practices, and what mentors are and are not positioned to do to share with stakeholders at participating sites.
- Establish practical routines for mentor and teacher communication about student math needs and progress, classroom curriculum, and difficulties observed during tutoring that may require additional support.
- Engage site leaders, school administrators, and teachers in setting expectations about protected tutoring time at the start of the year.

Develop mentors' math tutoring practice through targeted support

Consistent with Year 3 recommendations, the program expanded training and support for mentors in Year 4 by adding a lead mentor, distributing monthly resources through Google Classroom, and offering regional support. Mentors' assessments of onboarding and training were generally positive.

Year 4 findings point to the need to refine training and support in specific areas. Mentors were significantly more confident in relationship-building competencies than in tutoring competencies (a difference of 0.34 points), with the lowest ratings for using MLS to support learning and for helping students facing the greatest challenges. With well-established relational practices and experienced returning mentors in place, building mentors' knowledge and skills for the instructional tasks of tutoring can now be a focus for improvement. Mentors asked for opportunities to observe experienced mentors, guidance on structuring sessions, training on the MLS used at their sites, and strategies for behavior management, which focus group participants identified as the most common area for additional support.

Findings also point toward differentiating preparation by grade span. Students in eighth grade and above rated mentoring more than half a point lower than third graders, while one mentor working with elementary students described the Saga Coach training as geared toward older students.

Strategies to refine and improve mentor training and support may include the following.

- Refine mentors' training and ongoing support on tutoring competencies, including how to work with students with learning challenges.
- Differentiate training content by grade span, with attention to what building relationships and adapting instruction require in middle and high school settings, as compared to elementary settings.
- Create opportunities for mentors to observe and learn from peers within and across schools, including coaching opportunities between mentors.
- Coordinate with districts and schools on support that is best delivered locally, including math curriculum, student engagement, and behavior management practices aligned to local policies.

Strengthen program data infrastructure for monitoring, support, and evaluation

The participation log, designed and actively managed by the UEPC beginning this program year, produced monthly summaries of dosage by LEA, school, and mentor, and supported interim reporting for monitoring and continuous improvement. Additionally, participation rose across every data collection activity, which helps to identify strengths and areas for growth. At the same time, mentors and school staff described uncertainty and frustration with the data collection log, including mentors indicating that some tutoring sessions may not have been recorded. Clear guidance and timely feedback can both reduce the documentation burden and support the interpretation of the dosage data.

The remaining challenge is addressing the unevenness of coverage of data collection to ensure representativeness in the data across program sites and data sources. Low and/or non-participation in data collection, particularly from some sites with the highest concentrations of students identified as economically disadvantaged, leaves gaps in our understanding of how the program is working or could be improved based on varying contexts. Additionally, student survey response rates also ranged from 0.4% to 50% across LEAs, and two-thirds of school staff responses came from a single LEA.

To strengthen the usefulness of program and evaluation data, the following strategies may be helpful.

- Maintain the participation log and monthly reporting structure established in Year 4 while providing structured training on tutoring participation logs, and giving further guidance on which activities count as tutoring sessions and how to record them.
- Include participation log training in mentoring onboarding and provide feedback regularly to share progress on mentoring, inconsistencies, and areas that are being used to inform program improvement.
- Work with lead mentors at sites where participation in data collection has historically been lower to determine and address barriers to participation.

References

Altermatt, E. R., Altermatt, T. W., Rorrer, A. K., Groth, C., & Timmer, M. (2023). *STEM Action Center's AmeriCorps Math Mentors Program: 2022-2023 Evaluation Report.* Salt Lake City, UT: Utah Education Policy Center.

Altermatt, E. R., Gallyer, J., Altermatt, T. W., Yildiz, M., Rorrer, A. K., & Reynolds, A. L. (2026). *STEM Action Center's K-12 Math Personalized Learning Software Grant Program: 2025-2026 Evaluation Report.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2026/july/math_learning_software_evaluation_2026.php

Altermatt, E. R., Groth, C., & Rorrer, A. K. (2022*). AmeriCorps Math Mentors Program: 2021-2022 Pilot Year Report.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2022/november/math-mentors-program-2021-2022-pilot-year.php

Altermatt, E. R., Rorrer, A. K., & Moore, B. (2022). *Exploring Best Practices in Math Personalized Learning Software Use Among Educators Identified as Positive Outliers.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2022/november/math-software-best-practices.php

Altermatt, E. R., Rorrer, A. K., Altermatt, T. W., Doane, M., & Timmer, M. (2023). *Blended Learning Research Brief No. 2. Associations Between Student Attitudes Toward Math and Math Personalized Learning Software Use.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2023/september/blended-learning-research-brief-2.php

Altermatt, E. R., Yildiz, M., & Rorrer, A. K. (2024). *STEM AC's AmeriCorps Math Mentors Program: 2023-2024 Evaluation Report.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2024/august/americorp-math-mentors-program.php

Altermatt, E. R., Yildiz, M., & Rorrer, A. K. (2025a). *STEM AC's AmeriCorps Math Mentors Program: 2024-2025 Evaluation Report.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2025/august/americorps-math-mentors-report-2025.php

Altermatt, T. W., Altermatt, E. R., Rorrer, A. K., & Moore, B. (2022). *Math Personalized Learning Software: Examining Usage and Associations with Achievement in Utah During the Covid-19 Pandemic.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2022/november/math-personalized-learning-software-covid19.php

Betebenner, D. W., & Iwaarden, A. V. (2024). *Academic Impact and Recovery in Utah During the COVID-19 Pandemic.* Dover, New Hampshire: The National Center for the Improvement of Educational Assessment.

Bhatt, M. P., Chau, T., Condliffe, B., Davis, R., Grossman, J., Guryan, J., Ludwig, J., Magnaricotte, M., Mattera, S., Momeni, F., Oreopolous, P., & Stoddard. G. (2025). *Personalized Learning Initiative Interim Report: Findings from 2023-24*. University of Chicago Education Lab.

Bhatt, M., Guryan, J., Khan, S., LaForest-Tucker, M., & Mishra, B. (2024). Can technology facilitate scale? Evidence from a randomized evaluation of high dosage tutoring (NBER Working Paper No. 32510). National Bureau of Economic Research. https://doi.org/10.3386/w32510

Carbonari, M. V., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Kane, T. J., McDonald, A., McEachin, A., Morton, E., Muroga, A., Salazar, A., & Staiger, D. O. (2026). Impacts of academic recovery interventions on student achievement in 2022 to 2023. *Educational Evaluation and Policy Analysis, 48*(3), 1045–1072. https://doi.org/10.3102/01623737251400377

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cortes, K. E., Kortecamp, K., Loeb, S., & Robinson, C. D. (2025). A scalable approach to high-impact tutoring for young readers. Learning and Instruction, 95, 102021.

Crossley, S. A., Karumbaiah, S., Ocumpaugh, J., Labrum, M. J., & Baker, R. S. (2020). Predicting math identity through language and click-stream patterns in a blended learning mathematics program for elementary students. *Journal of Learning Analytics, 7*(1), 19–37.

Di Pietro, G., & Munoz, J. C. (2025). A meta-analysis on the effect of technology on the achievement of less advantaged students. *Computers & Education, 226*, 105197.

Eames, T., Brunskill, E., Yamkovenko, B., Weatherholtz, K., & Oreopoulos, P. (2026). Computer-assisted learning in the real world: How Khan Academy influences student math learning. Proceedings of the National Academy of Sciences, 123, e2507708123. https://doi.org/10.1073/pnas.2507708123

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. International *Journal of Behavioral Development, 31*(4), 317–327.

Guryan, J. Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R. G., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review, 113*(3), 738-765.

Guryan, J., & Ludwig, J. (2023). *Overcoming pandemic-induced learning loss.* In M. S. Kearney, J. Schardin, & L. Pardue (Eds.), Building a More Resilient US Economy (pp. 150 – 170). Washington, DC: Aspen Institute.

Hamlin, D., Peltier, C., & Reeder, S. (2026). Effects of a university-led high-impact tutoring program on low-achieving high school students: A 3-year randomized controlled trial. *American Educational Research Journal*. Advance online publication. https://doi.org/10.3102/00028312261450125

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computational Education, 153*.

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & de Boer, H. (2020). The relationships between school belonging and students' motivational, social-emotional, behavioural, and academic outcomes in secondary education: A meta-analytic review. *Research Papers in Education, 35*(6), 641–680. https://doi.org/10.1080/02671522.2019.1615116

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kraft, M. A., Sanderson Edwards, D., & Cannata, M. (2024). The Scaling Dynamics and Causal Effects of a District-Operated Tutoring Program. (EdWorkingPaper: 24-1030). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zcw7-4547

Kraft, M. A., Schueler, B. E., & Falken, G. (2024). *What Impacts Should We Expect from Tutoring at Scale? Exploring Meta-Analytic Generalizability.* (EdWorkingPaper: 24-1031). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zygj-m525

Lee, M. G., Loeb, S., & Robinson, C. D. (2024). *The Impact of High-Impact Tutoring on Student Attendance: Evidence from a State Initiative.* (EdWorkingPaper: 24 -1107). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/wghb-4864

Little, T. D., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Allen, P. J., & Noam, G. G. (2020). The retrospective pretest–posttest design redux: On its validity as an alternative to traditional pretest–posttest measurement. *International Journal of Behavioral Development*, *44*(2), 175–183. https://doi.org/10.1177/0165025419877973

Liu, J., Mbowe, A. E., Tahri, D., & Aziku, M. (2026). Meta-analysis on the influence of AI agents on K-12 student cognitive performance. *Computers in Human Behavior Reports, 21*, 100973. https://doi.org/10.1016/j.chbr.2026.100973

Markovitz, C. E., Hernandez, M. W., Hedberg, E. C., & Whitmore, H. W. (2022). Evaluating the effectiveness of a volunteer one-on-one tutoring model for early elementary reading intervention: A randomized controlled trial replication study. *American Educational Research Journal, 59*(4), 788–819. https://doi.org/10.3102/00028312211066848

McLeod, D. B. (1994). Research on affect and mathematics learning in the *JRME*: 1970 to the present. *Journal for Research in Mathematics Education, 25*(6), 637–647. https://doi.org/10.5951/jresematheduc.25.6.0637

Morton, E., Goldhaber, D., Kane, T. J., & McEachin, A. (2025). *School's in for summer: A scalable and effective post-pandemic academic intervention*. NWEA.

National Student Support Accelerator (2021). *Toolkit for Tutoring Programs*. Retrieved from https://doi.org/10.26300/5n7h-mh59

National Student Support Accelerator (2023*). Challenges and Solution to Implementing Tutoring at Scale.* Retrieved from https://nssa.stanford.edu/sites/default/files/Challenges%20Solutions%20Implementing%20Turoing%20at%20Scale.pdf

National Student Support Accelerator (2025). *Tutoring Quality Improvement System: Quality Standards. R*etrieved from https://nssa.stanford.edu/tqis/quality-standards

Nickow, A., Oreopoulos, P., & Quan, V. (2020). *The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence* (Working Paper No. 27476; Working Paper Series). National Bureau of Economic Research.

Norberg, K. A., Almoubayyed, H., De Ley, L., Murphy, A., Weldon, K., & Ritter, S. (2025). Rewriting content with GPT-4 to support emerging readers in adaptive mathematics software. *International Journal of Artificial Intelligence in Education, 35*, 587–626. https://doi.org/10.1007/s40593-024-00420-2

Pellegrini, M., Neitzel, A., Lake, C., & Slavin, R. (2021). Effective programs in elementary mathematics: A best-evidence synthesis. *AERA Open, 7*(1), 1-29.

Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring Program Outcomes: Using Retrospective Pretest Methodology. *American Journal of Evaluation*, *21*(3), 341–349. https://doi.org/10.1177/109821400002100305

de Ree, J., Maggioni, M. A., Paulle, B., Rossignoli, D., Ruijs, N., & Walentek, D. (2023). Closing the income-achievement gap? Experimental evidence from high-dosage tutoring in Dutch primary education. *Economics of Education Review, 94*, Article 102383. https://doi.org/10.1016/j.econedurev.2023.102383

Robinson, C. D., Gormley, D., & Loeb, S. (2025). *Five years of tutoring research: What we have learned since the pandemic.* https://nssa.stanford.edu/sites/default/files/Five%20Years%20of%20Tutoring%20Research%20What%20We%20Have%20Learned%20Since%20the%20Pandemic.pdf

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2024). *Design Principles for Accelerating Student Learning with High-Impact Tutoring* (Design Principles Series, Brief #30). Ed Research for Action.

Saldaña, J. (2025). *The coding manual for qualitative researchers* (5th ed.). SAGE.

Shmoys, R., McCormick, S., Bretas, S., Ready, D. D., McCarty, G., & Matulewicz, E. (2026). The dosage dilemma: Implementation evidence on the barriers to high-dosage tutoring. *Educational Researcher*. Advance online publication. https://doi.org/10.3102/0013189X261437761

University of Chicago Education Lab. (2025). *Catching up: Scaling high-impact tutoring in Miami-Dade*. https://educationlab.uchicago.edu/resources/catching-up-scaling-high-impact-tutoring-in-miami-dade/

Utah State Board of Education (USBE, 2026). *Annual Report 2026. https://schools.utah.gov/superintendentannualreport/_superintendentannualreport_/AnnualReport2026.pdf*

What Works Clearinghouse. (2017). *Beginning reading intervention report: Leveled literacy intervention*. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/WWC/InterventionReport/679

Yi, L., Liu, D., Jiang, T., & Xian, Y. (2025). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education, 23*(4), 1105–1126. https://doi.org/10.1007/s10763-024-10499-7

Young, J., & Kallemeyn, L. (2019). Testing the Retrospective Pretest with High School Youth in Out-of-School Time Programs. *Journal of Youth Development*, *14*(1), 216–229. https://doi.org/10.5195/jyd.2019.635

Appendices

Appendix A. Mentor Focus Groups Report (May 2026)

Interim Evaluation Report: Summary of Findings from Focus Groups with STEM Action Center Math Mentors

In the Spring of 2026, the Utah Education Policy Center (UEPC) conducted virtual focus groups with mentors from the Math Mentors Program to learn about their perspectives on the program. The UEPC facilitated **six virtual focus groups** between March 17 and March 23 with a total of **48 mentors** participating from across the state. Each session followed a semi-structured protocol, lasted approximately 50 minutes, and was audio-recorded and transcribed. Using a deductive analysis approach, the transcripts were coded to identify recurring themes related to three evaluation questions:

- How do mentors conceptualize their role in supporting students, and what strategies do they use to be effective in their role?
- How well does the mentor onboarding and training process prepare mentors, and what improvements could enhance its effectiveness?
- What are the key strengths and challenges of the mentoring program as identified by mentors, and what recommendations do they have for improvement at their site and program-wide?

Theme 1. Mentors viewed their roles in supporting students as both academic and relational

Math mentors have various backgrounds, such as retired teachers and school administrators, parents looking to use their math degrees within the schools their children attend, and college students exploring possible interests within public education. Regardless of their experiences that brought them to the program, the mentors all share a common purpose in helping students be successful in math through supplemental one-on-one or small group support during the school day by providing both relational and academic support.

Mentors explained that students were more willing to engage in math tasks after mentors established trust and personal connections with the students. Some mentors described how they built trust and connections through various informal interactions, noting that these were particularly important in helping students feel comfortable and supported. As one mentor described these types of informal interactions, she noted it was important to *"remember the small things"* and ask students about "*their day, hobbies, [and] other classes"* before working on the math tasks together. Another mentor explained how spending time observing and engaging with students during lunch and recess helped him to better understand his students across different settings, which he felt made him more effective as a math mentor.

Mentors also described the academic part of their role in providing math support for students who may need additional practice on math concepts, different ways to think about and solve math problems, and, for some students, a first-time experience of math success. For many mentors, this part

of their roles included incorporating math games, applying math vocabulary to students' real-world situations, and acknowledging mistakes as opportunities to learn and progress in their math skills.

Theme 2. Mentors viewed encouragement and confidence-building as central to students' success.

Mentors indicated that a central part of their role was to build students' confidence to take risks, make mistakes, and feel more comfortable performing math tasks independently. One mentor indicated that building trust was a precursor to confidence-building, noting that she tried to communicate to student that *"I know you can do this. I trust you to do this."* Another explained how she often tried to reframe the way students saw themselves, *"I'm so bad at math.' You look at what they're able to do and it's like, 'Oh, well, you're actually doing a great job.' So, trying to just help them reframe their mindset, be more positive."*

Mentor comments such as these reflect a broader pattern in which they viewed the relationships they'd built with their students as foundational so that they could then encourage them to persevere through difficult math tasks. For example:

- *"I've found it's really just about being there for the kids to have a shoulder…having someone just being right there to be able to give them confidence in their answers, and just even if they are wrong, just being able to put it on paper and do the work, first off. And I found that's a really big help when we're there with them. And when students see me coming in, they always are like, 'You're here finally. I can have someone to help me get through this math problem.' And it's awesome to see that kind of reaction just by being there."*
- *"I believe that my role as a math mentor is to give confidence and comfort in students as they're doing math. A lot of them are at least a little scared of math or being wrong. And so, I think I'm just there to make sure that they know they can do it themselves, even if they don't believe in it. And to get them comfortable that even when I'm not there, they can learn and do work on their own."*

Theme 3. Mentors conceptualized themselves as supplementing classroom instruction with additional math support and coordinating with teachers to determine students' specific math needs.

In addition to offering encouragement and building confidence among students, mentors also explained the other aspect of their roles as helping to develop and strengthen students' math skills. They described their roles in supplementing whole group math instruction with small group or one-on-one tutoring in an effort to increase students' interests, understanding, and success with math concepts and skills. One mentor gave an example of how she leveraged this supplemental support to help a student build critical math skills.

- *"I think we also are looking at deficiencies and we're trying to incorporate those into the learning so they became a strength instead of a deficiency. So at least for me, I analyze what they don't know and then try to figure out ways to teach that. I have several that struggle with multiplication and so we're constantly, when we have five minutes, we're doing flashcards or we're playing multiplication games or something to increase that. And I have a couple of them that have just grown leaps and bounds just by learning their multiplications."*

Additionally, some mentors explained how they intentionally coordinated with classroom teachers to identify specific needs of students who would benefit most from this supplemental support, as well as discuss formative progress on students who may need less support.

- *"I'll just talk to the teacher before class or just after class and we'll just kind of discuss how the students in my groups are doing. And some of them will be doing really good on a topic and they'll kind of just be phased out of my group. Then another one will just kind of be phased in just with how they're doing. And so just an open line of communication with those teachers of where the students are going."*

Comments such as these regarding ongoing communication with classroom teachers to coordinate specific math needs of students offers additional insight into how mentors view their roles of providing supplemental math instruction.

Theme 4. Mentors appreciated and valued the initial onboarding and periodic resources provided by the STEM Action Center.

Many mentors described the value of the initial onboarding provided by the STEM Action Center to prepare them for their new roles and responsibilities. During these onboarding sessions, mentors received context and background on the purpose of the program, instructions for submitting mentoring logs to document student participation, and other logistical information about the program. In addition to the initial onboarding sessions, mentors also appreciated resources shared periodically by the STEM Action Center (focused on topics such as questioning strategies, engagement ideas, and tips for building relationships with students).

- "*They've [STEM AC resources] been very great and they really help you reflect on things and trying to do better as a mentor and ways to encourage students. And even ways to look at math problems. And some of it has been super enlightening to see even how the math programs are being taught. And if they're ... I think I just did a training on rigor, so that one's in my head more, but what kinds of leveled questions they're even asking in math class and then encouraging us to do the same. So those are really great. And the best training, a lot of it I feel like is the on- the-job training. You just hop right in and you learn and as you go."*

Theme 5. Mentors identified specific opportunities for additional training, specifically aligned with school/district initiatives and policies, such as behavior management strategies, small group instruction, and math software training.

Although mentors generally felt supported and prepared for their new roles, they also identified opportunities to strengthen training and onboarding processes. Behavior management was one of the most common areas where mentors desired additional support, and they also acknowledged that this might be an area that schools and districts could provide training that is specific to their unique behavior programs or district policies. As one mentor explained:

- "*The only thing I struggle with sometimes is our school has a lot of behavioral issues, and I know that's not really the math mentor's job is to be the behavior controller, but it's maybe some tips and tricks for kids who are a little more behavioral. And we do get some of those trainings*

through our onboarding through the school, but a few more tips and tricks like that would be great because I always struggle with that."

Other mentors agreed with these opportunities for school/district trainings in behavior management, as well as in various instructional strategies that are aligned with district priorities or initiatives. As one mentor explained:

- "*We received really good training at our district level. We were able to get on and learn about small group work, learned about the 'big five'. And that was really helpful, I thought."* Finally, another mentor described how regional trainings are important for mentors instead of a "one-size-fits-all" approach, *"I think the onboarding gave me a good steppingstone, but I think it's almost when you get into a school, how it runs in ___ County and ___ County, even ___ County, they are totally different. Down here... I can't pull students out for a group. And this is one thing that has been a hard thing, is that I can't ... They're like, 'Pull them out for a group.' Then I would be pulling them from music, from library, from art. We don't have a lot of electives down here. So mainly the only time I can see a student one-on-one is during their actual math lesson or during their recess or their lunch, which then they're missing out on the social aspect of school. So, I think that that's a big thing*."

These mentor comments reflect a larger pattern suggesting that several onboarding topics, including specific MLS training, might be best taught by district or school teams that can speak to the local implementation rather than state-wide application.

Theme 6. Overall, mentors report that the program contributes to students' increased confidence and enjoyment of math, despite occasional school-specific program challenges.

Across focus groups, mentors identified several positive impacts of the program, such as increasing students' confidence in understanding and solving math problems, and willingness to engage in and persevere with challenging tasks. Additionally, some described watching their students learn to enjoy math through fun activities, real-life applications, and even experiencing success in math. As one mentor explained:

- "*That's the best part, I think is when they're like, 'Oh yeah, that's it.' And they get it, or they'll look at me like, 'Why didn't you tell me that before? This all makes sense now.' It's just that awesome feeling you get that they're starting to not just do math but embrace it and love it."*

Other mentors further elaborated on the program's strengths by considering the needs of students who benefited the most from this program as they explained:

- "*I feel like we're catching kids who would be lost if we weren't doing what we're doing."*

Theme 7. Mentors' recommendations for program improvements included having lead mentors at each school, creating opportunities for professional learning across schools, and ensuring clarity among school teams about math mentors' roles and responsibilities.

In responding to questions about program challenges and recommendations, some mentors described inconsistencies across schools in how mentors were utilized and supported. In some cases,

mentors felt that administrators and teachers were unclear about the mentor role or how mentors could best support classroom instruction. Others expressed that, while this was previously a problem at their school, it took a few years and consistent administrative support to develop clarity and understanding among the school team. The following comments from mentors explained the lack of clarity at many of the program sites:

- *"I think the biggest thing for me is a lot of the teachers and people at the school don't really quite understand what we're there for. Sometimes they just think, 'Oh, you're an extra aide,' or 'Oh, you're here to take the kid with behavioral issues for 30 minutes. Good, get them out of my classroom or something.' And really sit down with the teachers and be like, 'Look, these people are here to help your kids that are behind in math. They're here to help them improve their skills and build a relationship with them and stuff. They're not here to babysit. They're not here to walk around your classroom. This is what they're here for…"*
- *"They weren't sure what the difference between a math mentor was and an instructional assistant..."*

To strengthen the program moving forward, mentors recommended increasing opportunities for collaboration and communication among mentors, schools, and districts regarding the most impactful strategies that have generated the best math outcomes for students. Suggestions included establishing lead mentor positions at schools, developing opportunities for mentors to connect and learn from one another in professional learning communities, and providing clearer guidance related to mentor expectations and documentation processes. One mentor explained,

- *"For me, I think it would be interesting to see what other schools are doing, how they're doing this and just some consistency wise, because if all of a sudden you're looking at [school name] and [school name] is way out in the left field with how they're doing things, and then we try to compare that to see the value of the program."*

Conclusions

Overall, these findings suggest that mentors view the Math Mentor program as a valuable and impactful program that provides much needed support for students while also identifying opportunities to enhance consistency, onboarding/training, and program structures and processes across sites. Mentors across the state described meaningful relationships with students, moments of breakthrough in math understanding, and a genuine sense of purpose in their roles. These are all indicators of a program that is making a difference in students' lives. Their feedback also reflects the kind of thoughtful, constructive engagement that strengthens programs over time. The challenges mentors identified, including behavior management and ensuring that teachers and administrators fully understand the mentor role, are addressable with targeted support and clearer communication structures. Taken together, these findings offer both an affirmation of the program's core strengths and a roadmap for continuous improvement as the Math Mentor program continues to grow across Utah.

Appendix B. Interim evaluation report: Retrospective analysis of outcomes for the STEM Action Center's 2024-25 Math Mentors Program (February 2026)

Overview

The Utah STEM Action Center's ***Math Mentors Program*** provides small-group and one-to-one math tutoring to K–12 students, with the goals of strengthening student engagement and improving academic outcomes. During the 2024–25 school year, the program served approximately 2,000 students across 14 local education agencies (LEAs), including nine school districts and five charter schools.

In this interim report, the Utah Education Policy Center (UEPC) examines outcomes for participating students in 2024-25 using statewide administrative data from the Utah State Board of Education (USBE).[9] These analyses complement findings from the UEPC's 2024-25 annual report (Altermatt, Yildiz, & Rorrer, 2025), which relied on locally-provided data from eight of 14 participating LEAs. LEA-provided datasets offer certain advantages, including access to scores on the Acadience Math assessment for students in grades K–3. These scores are not available from the USBE. At the same time, analyses that rely on locally-provided data are subject to important limitations. Specifically, not all LEAs provide data, and demographic variables are often incomplete or inconsistently defined.



The statewide administrative data used in the current report provide a more complete and uniform view of program outcomes across LEAs. Importantly, these data also enable an initial examination of impacts on chronic absenteeism, which is particularly relevant given emerging evidence that high-dosage tutoring may improve student attendance (Lee et al., 2025).

Evaluation Questions

The primary goal of this interim report is to address the following evaluation questions:

> **EQ1.** Was participation in the Math Mentors Program in 2024-25 associated with improvements in academic achievement or school attendance?
> **EQ2.** Did program impacts vary by students' baseline academic performance and prior school attendance histories?

[9] The UEPC has a Master Data Sharing Agreement (MDSA) with the USBE permitting the use of education data for evaluation and research purposes. The UEPC adheres to the terms of the MDSA, including confidentiality, non-disclosure, data security, monitoring, and applicable state and federal laws and regulations. The UEPC also adheres to University of Utah Institutional Review Board provisions to protect student privacy.

Summary of Key Findings

Key findings from this report are summarized below. Detailed methods and results are presented in the sections that follow.

- *Math Mentors Program* participants did not differ significantly from similar non-participating students on end-of-year math achievement or attendance. However, the point estimate for chronic absenteeism favored participants, suggesting roughly 19% lower odds of chronic absenteeism among *Math Mentors Program* participants relative to matched peers. This corresponds to a small reduction of roughly three percentage points in the probability of chronic absenteeism (22% among participants compared to 25% among matched non-participants).
- Exploratory analyses indicated that, among *Math Mentors Program* participants without a prior history of chronic absenteeism, higher tutoring dosage was associated with lower predicted absenteeism rates.
- Tutoring dosage levels for most participants in the *Math Mentors Program* fell short of the approximately 900-minute minimum recommended by the National Student Support Accelerator (NSSA, 2025). Improving program dosage will be important for realizing the program's full potential impact on achievement and attendance outcomes.

Methods

To examine student outcomes associated with participation in the *Math Mentors Program*, tutoring dosage records from the STEM Action Center were linked to administrative records from the USBE. Because tutoring records did not include unique student identifiers, participation records were linked to USBE records using fuzzy matching procedures based on student names and enrollment information. USBE records included: enrollment data (school, grade level); demographics (gender, race/ethnicity, special education status, English Language Learner status, and low-income status); scores on end-of-year, statewide assessments of math achievement (RISE and Utah Aspire Plus); and indicators of chronic absenteeism.

In all, tutoring records for 1,773 students (or 83% of the 2,127 tutoring records available) were successfully matched to USBE administrative data. These matched records formed the starting point for all subsequent analyses.

Two primary outcomes were examined:

- **Math Achievement.** Math achievement was measured by students' scores on RISE (Grades 3-8) or Utah Aspire Plus (Grades 9-10) assessments. Scores were standardized by test type, grade level, and academic year to facilitate interpretation and comparison across grades. Scores from the 2024-25 school year served as the outcome measure, while scores from the prior school year were included as a baseline covariate. Analyses were limited to students in Grades 4–10, as standardized test scores were not available for students in Grades K–3 and were inconsistently available for students in Grades 11–12.
- **Chronic Absenteeism.** Attendance outcomes were measured using students' chronic absenteeism status. Chronic absenteeism was coded as a binary indicator reflecting whether a

student missed at least 10% of enrolled instructional days during the school year. Chronic absenteeism status from the prior school year was included as a baseline covariate.

Analytic Sample

The analytic sample was limited to students who received at least 90 minutes of tutoring during the academic year and who had complete data on demographic characteristics, baseline achievement, and prior attendance. After applying these criteria, 736 tutoring participants were eligible for inclusion in the analytic sample. Notably, tutoring dosage was a significant driver of sample attrition. Among the 1,773 students who were matched to the USBE dataset, only 1,123 (or 63%) received more than 90 minutes of tutoring during the academic year.



To estimate the association between *Math Mentors Program* participation and student outcomes while accounting for preexisting differences between participating and non-participating students, we used **propensity score matching** with nearest-neighbor 1:1 matching to construct a comparison group with similar baseline characteristics. Prior year academic achievement, prior year chronic absenteeism, English Language Learner status, and special education status were included to account for differences in student need and risk that may have influenced both selection into tutoring and subsequent outcomes. Grade level and district were included as exact matching variables to ensure that the two groups were exposed to similar academic standards and local implementation contexts.

Of the 736 eligible tutoring participants, 608 (83%) were successfully matched to comparable non-participating students, resulting in a final analytic sample of 608 tutoring participants and 608 matched non-participants. Balance diagnostics indicated strong baseline equivalence between the matched groups. As shown in Table 1, standardized mean differences (SMDs) for prior year math achievement, prior year chronic absenteeism, English Learner status, and special education status were small (all $|SMD| \leq 0.10$), indicating a high degree of similarity on observed baseline characteristics. We accounted for any remaining differences by including these and other covariates (e.g., gender) in all subsequent analyses.

Table 1. Post-Matching Balance Diagnostics for Propensity Score–Matched Sample

Covariate	Matched Participants	Matched Non-Participants	SMD (Post Match)
Prior year math achievement (z-score)	−0.513	−0.529	0.018
Prior year chronic absenteeism (%)	25.8	24.0	0.018
English Language Learner status (%)	14.5	13.3	0.012
Special education status (%)	15.5	20.9	−0.054

Note. Standardized mean differences (SMDs) reflect differences between matched participants and non-participants. Absolute values below 0.10 indicate good balance.

Analytic Approach

To evaluate the impact of the *Math Mentors Program* and to better understand patterns underlying student outcomes, we used a two-part analytic strategy. First, we estimated average program impacts by comparing participating students to the matched group of non-participating students. Second, recognizing that average impacts may obscure meaningful variation in students' experiences, we conducted exploratory analyses among tutoring participants to examine how outcomes varied with tutoring dosage and student characteristics.

To estimate **average program impacts**, we first used propensity score matching (as described above) to identify a comparison group of non-participating students with similar baseline characteristics to tutoring participants who received at least 90 minutes of tutoring during the academic year. We then estimated the average treatment effect for math achievement and chronic absenteeism using regression models applied to the matched sample.

To explore **heterogeneity in outcomes** among participating students, we conducted a second set of analyses restricted to tutoring participants only. These analyses examined the association between tutoring dosage (total minutes) and student outcomes, as well as whether these relationships differed by students' prior academic achievement or prior chronic absenteeism. Because tutoring dosage was not randomly assigned among participants, findings from these analyses should be interpreted as descriptive of variation in outcomes among tutored students and not as estimates of causal effects. To limit the influence of extreme values, participants (n = 10) who received more than 1,800 minutes (or 30 hours) of tutoring were excluded from these analyses.

Findings

Average Program Impacts

Across the propensity score–matched sample, we did not observe statistically significant average differences between tutoring participants and matched non-participants on either outcome. For math achievement, the estimated average treatment effect was small in magnitude (approximately −0.04 standard deviations) and not statistically significant ($p > .05$), indicating no detectable difference in end-of-year achievement between groups after accounting for prior achievement and student characteristics. Attendance outcomes told a more encouraging story. Here, the point estimate was favorable for *Math Mentors Program* participants, suggesting approximately 19% lower odds of being chronically absent relative to matched non-participants. This corresponds to a small reduction of roughly three percentage points in the probability of chronic absenteeism (22% among participants compared to 25% among non-participants). However, this difference was not statistically significant ($p > .05$) and should be interpreted with caution.

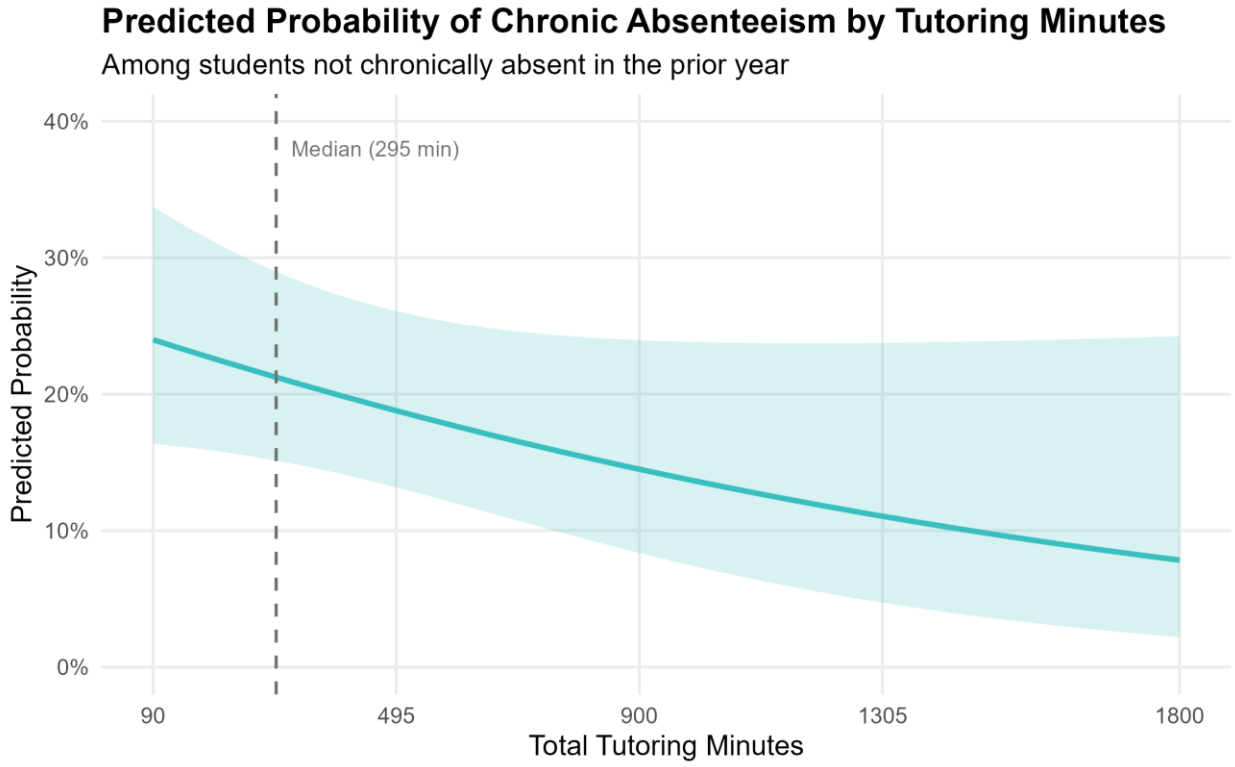
Heterogeneity in Achievement Outcomes Among Participants

Among tutoring participants, a significant interaction was observed between tutoring dosage and prior achievement ($p < .05$), indicating that the relationship between minutes of tutoring received and end-of-year math scores varied by students' incoming performance level. The direction of findings suggested that students with greater academic need may have derived more benefit from higher dosage, while students closer to average achievement levels in the prior year showed little gain or

modestly lower scores at higher dosage levels. However, predicted differences across dosage levels were small and confidence intervals overlapped substantially across groups, indicating limited practical significance. These patterns should be interpreted cautiously, as students who received more tutoring minutes may differ systematically from those who received fewer in ways not fully captured by the model.

Heterogeneity in Attendance Outcomes Among Participants

Among tutoring participants, a significant interaction was also observed between tutoring dosage and prior chronic absenteeism status ($p < .05$), indicating that the relationship between minutes of tutoring received and attendance varied by students' incoming chronic absenteeism status. For tutoring participants who were not chronically absent in the prior year (76% of the sample), greater tutoring dosage was associated with somewhat lower predicted probabilities of chronic absenteeism during the 2024–25 school year. As shown in the figure, predicted chronic absenteeism declined as tutoring minutes increased, from about 24% at 90 minutes to about 8% at 1,800 minutes. Importantly, most students in both subgroups (i.e., those with and without prior chronic absenteeism) received relatively low levels of tutoring. The median dosage for the group of students who not chronically absent in the prior year was 295 minutes, a level associated with only modest reductions in predicted chronic absenteeism. This suggests that many participants may not have received sufficient tutoring to fully realize the program's potential attendance benefits.



Note. Predicted probabilities are averaged over observed values of covariates and restricted to students not chronically absent in the prior year. The dashed vertical line indicates the median tutoring dosage among this group (295 minutes). Confidence bands reflect 95% confidence intervals and widen at higher dosage levels where data are sparse.

A different trend emerged for students with a history of chronic absenteeism in the prior year. For these students, higher tutoring dosage was associated with somewhat higher predicted probabilities of chronic absenteeism. This pattern is likely driven by selection into dosage levels, with students facing more severe or persistent attendance challenges receiving more intensive tutoring, rather than indicating that increased tutoring worsens attendance. As with the achievement analyses, these results should be interpreted cautiously given the observational nature of tutoring dosage and the absence of random assignment within the participant sample.

Summary and Implications

This interim report examined outcomes associated with participation in the Utah STEM Action Center's *Math Mentors Program* using statewide administrative data linked to tutoring records from the 2024–25 school year. Findings were mixed but informative.

On average, tutoring participants did not differ significantly from matched non-participants on end-of-year math achievement or chronic absenteeism. The absence of statistically significant average effects should be interpreted in context. Although a significant body of research highlights the potential for high-dosage tutoring to have significant positive benefits on student learning outcomes (Accelerate, 2026; Guryan et al., 2023; Robinson, Kraft, Loeb, & Schueler, 2024), demonstrating measurable impacts for high-dosage tutoring programs at scale is well documented as challenging. For example, a 2024 meta-analysis of randomized controlled trials found that tutoring effects tend to be smaller – and in some cases negative – in programs serving more than 1,000 students, where dosage often declines, student–tutor ratios increase, and implementation quality becomes more difficult to maintain (Kraft, Schueler, & Falken, 2024; see also Kraft, 2020; Kraft, Sanderson Edwards, & Cannata, 2024). Together, this body of research suggests that null average impacts in large-scale tutoring evaluations are not uncommon and may reflect implementation constraints rather than a lack of program promise.

Consistent with this interpretation, exploratory analyses among tutoring participants revealed two notable patterns. First, the relationship between tutoring dosage and math achievement appeared to vary by students' prior performance, with higher-need students showing more favorable trends at higher dosage levels. Second, and more compellingly, higher tutoring dosage was associated with substantially lower predicted probabilities of chronic absenteeism among students without a prior history of attendance problems. Notably, the median tutoring dosage among these participants was 295 minutes—well below the range in which the most pronounced attendance benefits are predicted. This pattern suggests that limited program intensity may have constrained observable impacts and that increasing dosage represents a promising avenue for maximizing program effectiveness.

Taken together, these findings suggest that the *Math Mentors Program* shows promise, particularly with respect to attendance outcomes, but that current dosage levels may be insufficient to produce detectable average effects. Future analyses should examine whether impacts strengthen as program implementation matures and as more students accumulate higher dosage levels.

References

Accelerate (2026, February). *Accelerate's 2024–25 Call to Effective Action: A Synthesis of Lessons Learned* (PDF). https://accelerate.us/wp-content/uploads/2026/02/Accelerate-CEA_Synthesis-Report_24-25-FINAL-020826-1.pdf

Altermatt, E. R., Yildiz, M., & Rorrer, A. K. (2025). *STEM Action Center's AmeriCorps Math Mentors Program: 2024-2025 Evaluation Report.* Salt Lake City, UT: Utah Education Policy Center. https://uepc.utah.edu/our-work/posts/2025/august/americorps-math-mentors-report-2025.php

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R. G., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review, 113*(3), 738-765.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241-253.

Kraft, M. A., Schueler, B. E., & Falken, G. (2024). *What Impacts Should We Expect from Tutoring at Scale? Exploring Meta-Analytic Generalizability.* (EdWorkingPaper: 24-1031). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zygj-m525

Kraft, M. A., Sanderson Edwards, D., & Cannata, M. (2024). The Scaling Dynamics and Causal Effects of a District-Operated Tutoring Program. (EdWorkingPaper: 24-1030). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zcw7-4547

Lee, M. G., Loeb, S., & Robinson, C. D. (2025). The Impact of High-Impact Tutoring on Student Attendance: Evidence from a State Initiative. (EdWorkingPaper: 24-1107). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/wghb-4864

National Student Support Accelerator (NSSA). (2025). *Tutoring quality standards*. Stanford University. https://nssa.stanford.edu/tqis/quality-standards

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2024). *Design Principles for Accelerating Student Learning with High-Impact Tutoring* (Design Principles Series, Brief #30). Ed Research for Action.

Appendix C. Supplementary tables for student surveys

C1. Scale reliabilities for student survey measures

Scale	Time	Items	N	Cronbach's α
Confidence	Retrospective-pre	3	698	0.845
	Post	3	697	0.814
Effort	Retrospective-pre	3	698	0.843
	Post	3	697	0.774
Interest/Value/Identity	Retrospective-pre	3	694	0.774
	Post	3	694	0.742
Growth Mindset	Retrospective-pre	3	696	0.829
	Post	3	693	0.847
Satisfaction	Post	3	696	0.819
Software	Post	3	552	0.869

Note. α = Cronbach's alpha, a measure of internal consistency, computed separately for the retrospective pre- and post-measures. All scales exceed the conventional .70 threshold at both timepoints.

C2. Scale and item-level student survey results summary

	Retrospective Pre-			Post-		
Scale/Item	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**
Math Confidence	**3.06**	**1.09**	**698**	**3.57**	**0.92**	**697**
How confident are you that you can complete all the work that is assigned in math?	3.07	1.26	695	3.70	1.03	696
When complicated ideas are presented in math, how confident are you that you can understand them?	2.96	1.25	696	3.32	1.10	691
How confident are you that you can learn all the material presented in math?	3.14	1.21	691	3.70	1.10	685
Math Effort	**3.83**	**0.97**	**698**	**4.21**	**0.76**	**697**
How hard do you try in math?	3.79	1.14	694	4.25	0.89	695
How hard do you try to pay attention in math?	3.80	1.11	693	4.12	0.95	693
How hard do you try to learn the math you are taught?	3.90	1.08	692	4.28	0.88	689
Math Interesting, Valuing, and Identity	**3.31**	**1.08**	**694**	**3.70**	**0.95**	**694**
How interesting do you find the things you learn in math?	3.03	1.31	693	3.37	1.27	693
How important is it to you to do well in math?	3.79	1.25	692	4.24	1.00	689
How much do you see yourself as a person who does well in math?	3.10	1.33	690	3.49	1.21	687
Growth Mindset	**3.70**	**1.03**	**696**	**4.16**	**0.89**	**693**
How much do you believe that a person's math skills can get better with hard work?	3.78	1.18	694	4.29	0.97	692
How much do you believe that you can get better at math if you keep trying?	3.79	1.17	694	4.29	0.97	691
How much do you believe you can learn math even when it feels hard?	3.53	1.23	691	3.90	1.11	686

Scale/Item	Retrospective Pre-			Post-		
	Mean	SD	N	Mean	SD	N
Mentor Satisfaction	N/A, post-only			**4.37**	**0.79**	**696**
How much did you like learning from your mentor?				4.41	0.90	694
How much did your mentor help you improve your skills in math?				4.39	0.89	693
How much did you mentor help you improve your confidence in math?				4.31	0.96	690
MLS Value	N/A, post-only			**3.92**	**1.04**	**552**
How much do you like using the math computer program with your mentor?				3.83	1.20	532
How much does using the math computer program with your mentor help you improve your skills in math?				3.92	1.13	538
How much does using the math computer program with your mentor help you improve your confidence in math?				3.95	1.17	535

C3. Grade-level regression models for student survey outcomes

	Predicted Change Scores for Outcomes				Post-Only Valuing	
	(1) Confidence	(2) Effort	(3) IVIM	(4) Growth Mindset	(5) Mentoring Satisfaction	(6) MLS Satisfaction
3rd	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
4th	0.234	0.170	0.197	0.198	-0.104	-0.159
	(-0.21)	(-0.196)	(-0.199)	(-0.257)	(-0.078)	(-0.119)
5th	0.220*	0.227	0.164	0.306	0.066	-0.312*
	(-0.102)	(-0.123)	(-0.139)	(-0.186)	(-0.089)	(-0.139)
6th	0.083	-0.019	0.130	0.156	-0.153	-0.344**
	(-0.22)	(-0.129)	(-0.111)	(-0.146)	(-0.115)	(-0.116)
7th	0.073	0.002	0.002	0.148	-0.280*	-0.334*
	(-0.129)	-(0.099)	(-0.092)	(-0.125)	(-0.104)	(-0.12)
8th	0.0218	0.179	0.164	0.379*	-0.576***	-0.919***
	(-0.191)	(-0.151)	(-0.123)	(-0.136)	(-0.075)	(-0.224)
9th - 12th	0.220	0.232*	0.032	0.375**	-0.683***	-0.915***
	(-0.123)	(-0.083)	(-0.083)	(-0.119)	(-0.079)	(-0.107)
Constant	0.502***	0.417**	0.338**	0.445*	4.457***	3.969***
	(-0.093)	(-0.119)	(-0.107)	(-0.153)	(-0.126)	(-0.106)
Observations	687	686	683	683	688	545
Adjusted R-squared	0.019	0.033	-0.001	0.019	0.047	0.069

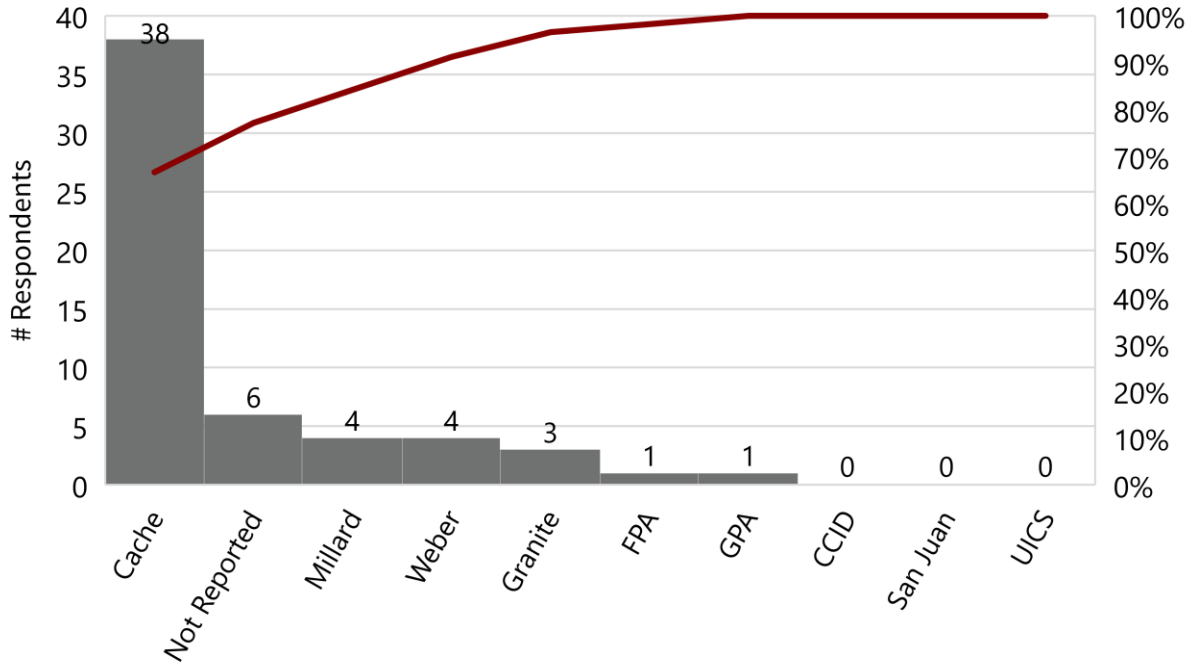
** p<0.05, ** p<0.01, *** p<0.001*

Note. Estimates are adjusted for LEA, and all standard errors are clustered at the school level.

Appendix D. School staff survey results summary

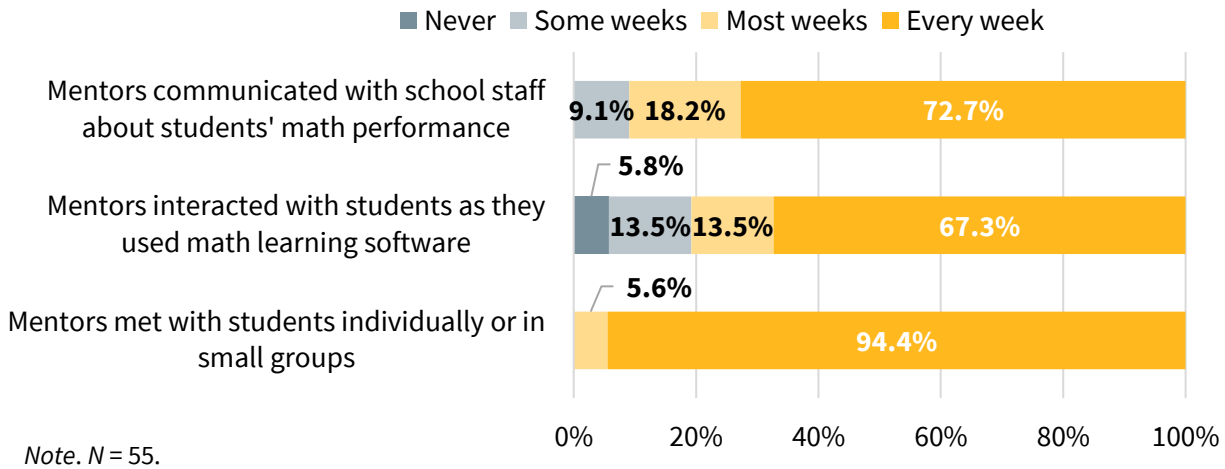
The results of the MM School Staff survey results are reported separately as they are not representative of LEAs and schools across the program. As seen in Figure D1, below, two-thirds of respondents came from just one LEA, with just one to four responses from other LEAs. A summary of the results of this survey are provided below (Figures D1-D4), though we caution they may not accurately represent the perspectives of MM school staff across LEAs.

Figure D1. Concentration of MM School Staff Survey Respondents by LEA



Note. *N* = 57.
Source: School Staff EOY Survey

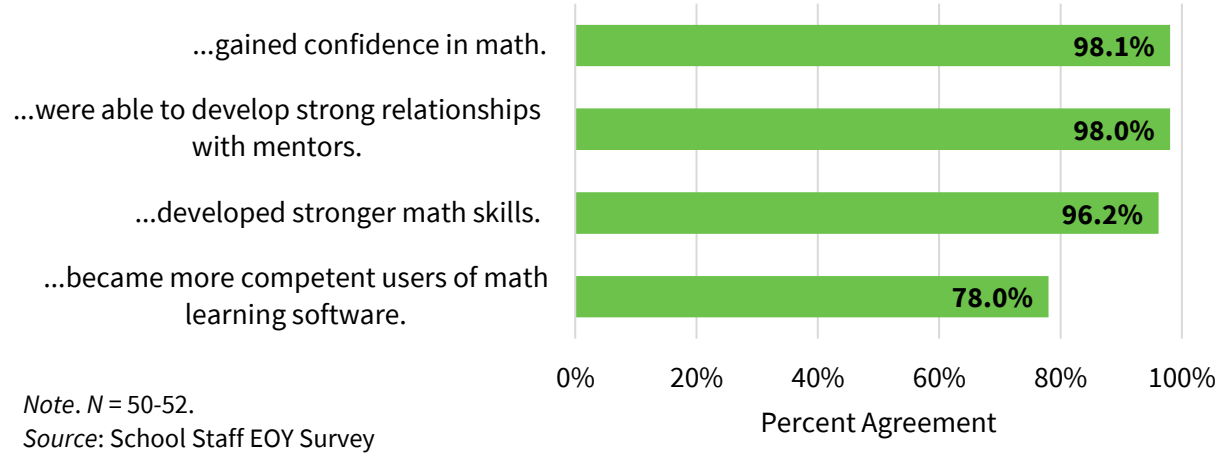
Figure D2. Implementation: MM School Staff Reports of Mentors' Tutoring Practices



Note. *N* = 55.
Source: School Staff EOY Survey

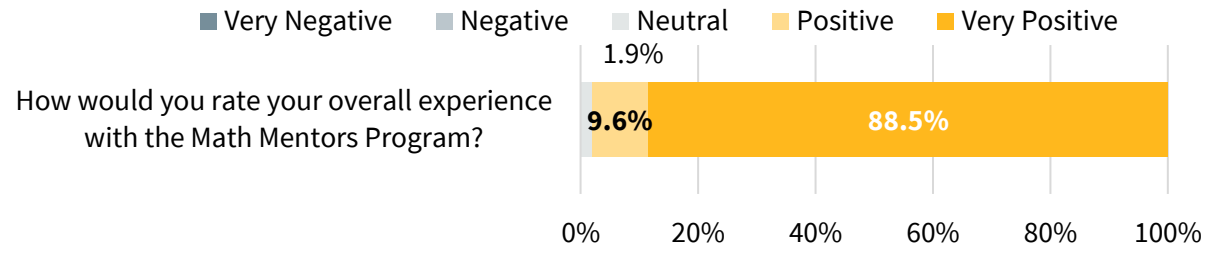
Figure D3. MM School Staff Perceptions of the Impacts of the MM Program on Students

As a result of participating in the STEM AC MM program, students...



Note. *N* = 50-52.
Source: School Staff EOY Survey

Figure D4. MM School Staff Overall Experience With the Program



Note. *N* = 52.
Source: School Staff EOY Survey

Appendix E. Sensitivity analyses for achievement outcomes

To assess the robustness of the preliminary achievement findings reported in the student achievement outcome evaluation section, we conducted three additional analyses using alternative definitions of tutoring exposure and student achievement. Results are summarized in Table F1.

For LEA 1, all three sensitivity analyses produced statistically significant positive estimates. The continuous-dosage model was positive (estimate = 0.010, $p = .031$), as was the residualized-growth model (estimate = 0.264, $p = .001$). In the propensity-score-matched analysis, students receiving at least 450 minutes of tutoring also had higher achievement than matched non-participants (estimate = 0.299, 95% CI [0.099, 0.499], $p = .004$).

Results for LEA 2 were similarly consistent. Continuous dosage was positively associated with achievement (estimate = 0.016, $p = .042$), the residualized-growth analysis produced a positive estimate (estimate = 0.334, $p < .001$), and the propensity-score-matched comparison favored students receiving 450 or more minutes of tutoring (estimate = 0.404, 95% CI [0.215, 0.592], $p < .001$).

For LEA 3, all three sensitivity analyses produced negative estimates. Continuous tutoring dosage was negatively associated with achievement; the residualized growth model produced a negative participation estimate; and the matched comparison of 450+ minute students with non-participants also produced a negative estimate. Additional analyses conducted separately for each of LEA 3's schools likewise did not identify evidence of a positive association between MM tutoring and mathematics achievement. **Thus, the absence of a positive result in LEA 3 does not appear to be driven solely by pooling the schools or by one particular model specification.**

The consistency of these results across alternative specifications strengthens the overall pattern emerging from the preliminary analysis. Although the magnitude of individual estimates varies according to the analytic approach, **LEA 1 and LEA 2 consistently show evidence of a positive association between MM tutoring and mathematics achievement, whereas LEA 3 consistently does not.**

Table E1. Sensitivity Analyses of the Association Between MM Tutoring and Math Achievement, by LEA

LEA	Continuous Dosage	Residualized Growth	PSM: 450+ Min. vs. Non-Participant
LEA 1	0.010 [0.001, 0.019], $p = 0.031$	0.264 [0.108, 0.420], $p = 0.001$	0.299 [0.099, 0.499], $p = 0.004$
LEA 2	0.016 [0.001, 0.032], $p = 0.042$	0.334 [0.199, 0.469], $p = <.001$	0.404 [0.215, 0.592], $p = <.001$
LEA 3	-0.009 [-0.011, -0.006], $p = <.001$	-0.179 [-0.228, -0.130], $p = <.001$	-0.115 [-0.196, -0.034], $p = 0.005$

Note: Sensitivity analyses include models using continuous tutoring dosage, residualized achievement growth, and propensity score matching comparing students receiving 450+ minutes of tutoring with matched non-participants. Because the analyses use different measures of tutoring exposure and/or achievement, coefficients should not be directly compared across columns.
Sources: Participation logs and LEA student achievement and enrollment data

Project Staff

The following Utah Education Policy Center (UEPC) team members contributed to this project.

Amy L. Reynolds, Ph.D.
Senior Research and Evaluation Associate

Ellen Altermatt, Ph.D.
Research and Evaluation Consultant

Muhammed Yildiz, Ph.D.
Research and Evaluation Consultant

Andrea K. Rorrer, Ph.D.
UEPC Director

Sarah R. Morris, Ph.D.
Research and Evaluation Associate

Lisa Wisham, M.Ed.
Senior School Support and Research Associate and School and District Improvement Specialist

James Gallyer, Ph.D.
Senior Data Scientist

Matt Doane, MPP
Senior Research and Evaluation Associate