Mathematica®
Progress Together

Blueprint Math Fellows Tutoring Program: Math Knowledge Impacts and Participant Math Perceptions

Andrew Gothro and Greg Chojnacki

March 2023

Contents

Tables

Figures

Middle Years Math Grantee Report Series

This report is one in a series of six reports on math tutoring programs. Over the 2020–2021 and 2021–2022 school years, the Bill & Melinda Gates Foundation invested in rapid-cycle evaluations of a cohort of 10 tutoring providers to learn about their innovative approaches to tutoring as part of its Middle Years Math body of work.[1] The goal of these investments was to understand how different tutoring models might create positive student experiences and lead to improved academic outcomes for students in the foundation's priority communities—those who are Black, Latino, and/or experiencing poverty. These investments were grounded in the substantial body of evidence supporting the effectiveness of tutoring in improving student math knowledge (Nickow et al., 2020).

To build on this existing evidence of effectiveness, the Gates Foundation sought to develop new early evidence about the success of a range of tutoring approaches. Specifically, these investments targeted two key learning priorities. First, the foundation sought to learn how innovative technologies and tutoring program design features might simultaneously improve the quality and lower the cost of tutoring, making high-quality tutoring available to a large number of students in priority communities. The second priority was to learn the extent to which tutoring programs resulted in positive experiences for participating students. To learn about tutoring design features, the foundation invested in tutoring programs with a wide range of approaches, including group and one-on-one tutoring, virtual and in-person models, professional teachers as tutors, or volunteer tutors who shared aspects of identity with tutored students. Tutoring programs also used different approaches to tutoring curriculum and pedagogy. The goal of this report series is to inform the tutoring field more broadly and support the provision of high-quality tutoring to as many students in the priority communities as possible.

To learn rapidly about tutoring providers' innovative approaches, Mathematica worked with each one to identify the most rigorous study design that would be feasible for district partners within a one-to-three-month planning period. Some providers were able to design and implement randomized controlled trials; others used quasi-experimental designs such as matched comparison approaches. One study compared growth in math knowledge among participants to the growth observed in national samples because it was not possible to obtain student-level data for comparison students who did not receive tutoring. These relatively small studies were right-sized to the development stage of the tutoring program and sought to demonstrate early evidence of success before moving on to larger-scale effectiveness studies. To help synthesize findings about student experiences from multiple providers, studies used the same student survey measures of tutor relationship, math confidence, and sense of belonging in tutoring sessions. Most of the studies used standardized math knowledge assessments aligned with Common Core State Standards.

Each study also aimed to inform providers' efforts to refine their programs and support successful implementation. These studies measured the amount of tutoring offered, attendance, and staff impressions about implementation challenges while also gathering qualitative data on students' experiences. Findings from these studies have helped to direct tutoring providers' next steps in refining and scaling their tutoring programs.

[1] This publication is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Executive Summary

What is the tutoring program we studied?

The Blueprint Math Fellows program provided tutoring and a web-based math learning platform to middle school students during their scheduled school day, five days per week in 45-minute sessions. Blueprint's Math Fellows program includes four key components: in-person and online live group tutoring, use of the web-based math learning platform ALEKS (Assessment and LEarning in Knowledge Spaces), family engagement, and student incentives. For instruction, students rotated between two in-person tutors and a remote tutor, in groups of two to 10 students. Blueprint regrouped students regularly based on data on students' performance. However, Blueprint assigned students to a consistent in-person or live remote tutor for "homebase" sessions. Homebase group tutoring sessions occurred at least twice per week and included instructional time, number talks, independent work on ALEKS, and goal setting. Blueprint assigned students to different sized homebase groups to test effects of group size on student outcomes. Blueprint tutors, called "fellows," are AmeriCorps volunteers who receive training, support, and an hourly stipend from Blueprint to provide structured math tutoring. See Appendix A for additional details.

What questions does this study answer?

1. Among students identified to receive the tutoring program, what is the average attendance rate? Does attendance vary by size of the homebase tutoring group?
2. Do students who participate in the tutoring program report having a high-quality relationship with their tutors? Does the reported quality of students' relationships vary by size of the homebase tutoring group?
3. Do students who participate in the tutoring program score higher on district assessments than students who do not participate in the tutoring program? Does this difference vary by student characteristics or size of the homebase tutoring group?
4. Do students who participate in the tutoring program report higher levels of math confidence and sense of belonging after receiving tutoring than before? Does this vary by size of the homebase tutoring group?

How was the study conducted?

Study design. For the 2021–2022 school year, school administrators at two sites in one suburban district assigned 185 students in grades 6, 7, and 8 to Blueprint tutoring. Administrators selected students primarily based on whether they had available intervention periods aligned with Blueprint's program schedule. Students in the same school and grade who were not assigned to Blueprint tutoring formed the comparison group (with a baseline sample of 1,088), and they received other elective or intervention instruction in its place. Most students in the analysis sample (88 percent) were Hispanic. (We use the term Hispanic to reflect the demographic data available in the participating district.) Blueprint tutoring did not replace core math instruction in students' schedules. Students participated in Blueprint for the entire school year. When program vacancies appeared from students transferring out of the school, school administrators assigned additional students to the program. To explore

students' experiences in different group sizes, at the beginning of the school year, Blueprint randomly assigned students to large (eight students) or small (four students) homebase groups. During the year, group sizes that students experienced changed due to student turnover. The goals of this study were to learn about Blueprint tutoring's effects on students' knowledge and perceptions of math as well as how those outcomes varied across different sizes of homebase tutoring groups.

Measures and analysis. Blueprint administered surveys to students participating in the program at the start of the program in September 2021 and follow-up surveys in May 2022. Baseline surveys collected information about students' confidence in math and sense of belonging in tutoring and in math class. Follow-up surveys repeated those measures and collected information about students' relationships with their Blueprint tutors. To assess effects on student math knowledge, the study team analyzed data from Star Math assessments, which the district administered in fall 2021 and spring 2022 to all students, including those receiving Blueprint tutoring and those in the comparison group. We used a combination of analytic methods to answer the research questions. The study team used descriptive methods to examine student attendance in tutoring sessions, student–tutor relationships, and student enjoyment of math. We used a Bayesian regression-adjusted comparison group design to compare math performance of students who participated in tutoring with that of students in the same district who did not participate in tutoring. To complement the main analyses, we also used descriptive methods to examine the relationships between size of the homebase tutoring group, student attendance, and student math performance. Due to fluctuations in homebase groups during the year, we analyzed outcomes for three sets of students based on group sizes that students experienced at the end of the school year: small (two to four students), medium (five to seven students), and large (eight to 10 students).

Limitations. Relatively low survey response rates (52.3 percent at baseline and 62.2 percent at follow-up) might have affected the accuracy of the estimated results. This can affect the results if students who did not respond to the survey differ from those who did in ways related to study outcomes. Additionally, Blueprint students scored higher on the baseline assessment than comparison group students. Although the study team adjusted for differences in baseline test scores and student characteristics in the analysis, these statistical adjustments may not fully account for all differences between the two groups.

What did the study find?

Attendance in tutoring. Student attendance in tutoring sessions averaged 81.5 percent during the school year. Attendance varied only slightly across students in homebase groups of different sizes.

Student-tutor relationships. Most students (76.8 percent) reported strong relationships with their tutors. We considered reported relationships to be strong when students' responses averaged 4 or higher on a 5-point scale. A much higher share of students reported strong relationships in Site 2 (92.3 percent) than Site 1 (55.3 percent). A smaller share of students ending the year in large homebase groups (70.3 percent) reported strong relationships with

their tutors, compared to students ending in small groups (77.3 percent) or medium-sized groups (83.9 percent).

Math knowledge. End-of-year Star Math assessment scores for Blueprint participants were 0.12 standard deviations higher than those of their nonparticipant peers, after controlling for differences in baseline scores and characteristics. Based on this estimate, there is a 99 percent probability that the program improved student scores. Results varied by site, with a larger effect in Site 1 (0.14 standard deviations) than in Site 2 (0.09 standard deviations), and they also varied by student characteristics. Descriptively, students in large homebase groups showed greater improvement in Star Math scores than those in either medium or small groups.

Student math confidence and sense of belonging. Student confidence in math and sense of belonging in tutoring increased modestly during the school year. Responses varied across sites, however, with reported confidence and sense of belonging in tutoring decreasing slightly for students in Site 1 and increasing for students in Site 2. Students ending the year in medium-sized homebase groups reported a meaningful improvement in sense of belonging in tutoring (an increase of 0.8 points on a 5-point scale), compared to no substantial change for students in other group sizes.

This study provides further evidence of the potential benefits of tutoring on student math knowledge, along with suggestive evidence on how program variations might affect outcomes. In particular, the results suggest that larger tutoring groups—as high as eight or 10 students—might still lead to positive math knowledge gains, at least in an in-school tutoring model implemented by a seasoned tutoring organization. Results suggest smaller groups might be more conducive to forming strong student–tutor relationships, but those results did not correlate with larger improvements in math knowledge scores in the Blueprint tutoring program.

Introduction

Blueprint Schools Network is a nonprofit organization providing educational services to schools and students. The goal of Blueprint's hybrid math tutoring program is to ensure that all students receive the math instruction necessary to rise to their full academic and social potential. Blueprint aims to level the playing field so that every student, regardless of race and family income status, has access to high-quality instruction and academic support. The Blueprint tutoring program does this by drawing on the strengths of in-person and online tutoring and web-based math learning platforms, and by providing incentives to for student participation and family engagement. Blueprint tutors are trained AmeriCorps volunteers who receive a stipend in exchange for providing tutoring.

In the 2021–2022 school year, Blueprint tested the effects of a pilot project that varied several elements of Blueprint's traditional tutoring program. Before the COVID-19 pandemic, Blueprint offered exclusively in-person tutoring. During the 2020–2021 school year, the two middle schools serving as the study setting operated primarily virtually, and Blueprint shifted to primarily virtual live tutoring. In 2021–2022, the time frame covered by this analysis, Blueprint piloted a hybrid approach, combining in-person and virtual live tutoring. Two in-person tutors and a live remote tutor shared instructional responsibility for an entire class (20–25 students). The composition of student instructional groups changed regularly based on performance data from ALEKS and interim assessments. In-person tutors delivered lessons that aligned generally with each school's scope and sequence. The remote tutor provided online, individualized support to students who were working independently on ALEKS topics during class. Although instructional groupings were fluid, students were also assigned to one of these three tutors for the entire school year for additional activities. Twice weekly, this "homebase" tutor conducted number talks and goal-setting, team-building, and motivational activities with students in one of three group sizes: large (eight to 10 students), medium (five to seven students), or small (two to four students).

The goal of this study is to measure the effects of Blueprint's hybrid tutoring program on math knowledge, student–tutor relationships, and student sense of belonging and confidence in math for middle school students in two schools. Analyses highlight differences in outcomes across the two sites. A secondary descriptive analysis compares outcomes across students ending the year in small, medium, or large homebase tutoring groups. We address the following research questions:

1. Among students identified to receive the tutoring program, what is the average attendance rate? Does attendance vary by size of the homebase tutoring group?
2. Do students who participate in the tutoring program report having a high-quality relationship with their tutors? Does the reported quality of students' relationships vary by size of the homebase tutoring group?
3. Do students who participate in the tutoring program score higher on district assessments than students who do not participate in the tutoring program? Does this difference vary by student characteristics or size of the homebase tutoring group?
4. Do students who participate in the tutoring program report higher levels of math confidence and sense of belonging after receiving tutoring than before? Does this vary by size of the homebase tutoring group?

Summary of Findings

Findings from the study of Blueprint's tutoring program in two predominantly Hispanic suburban middle schools in the 2021–2022 school year were positive but varied across sites. Participating in the program was associated with a moderate positive effect on student math test scores, with a larger effect in Site 1 than Site 2. By contrast, student survey outcomes were more positive in Site 2 than Site 1, including stronger reported student–tutor relationships and increases in reported math confidence and sense of belonging in tutoring. Table 1 shows the findings overall and by site. These differences might reflect differences in the instructional environment between the two sites—Site 1 experienced more instructional disruption due to COVID-19 than Site 2 did, for example—potentially creating a stronger contrast between Blueprint and non-Blueprint students. Finally, ending the year in large homebase groups did not appear to result in decreased growth in math knowledge, although those in large groups reported weaker student–tutor relationships.

Table 1. Student outcomes overall and by site

Outcome	Overall	Site 1	Site 2
Percentage of tutoring sessions attended	81.5	79.5	88.3
Effect on Star Math (standard deviations)	0.12	0.14	0.09
Percentage reporting a strong relationship with tutor (spring survey)	76.8	55.3	92.3
Change in reported sense of belonging in tutoring (on a scale of 1–5)	0.3	-0.2	0.7
Change in reported confidence in math (on a scale of 1–5)	0.1	-0.1	0.2

Source: District and Blueprint administrative data and student survey data.

Key Findings

Attendance in tutoring sessions was high.

Student attendance in tutoring sessions averaged 81.5 percent during the school year, slightly trailing overall school attendance among Blueprint students (86.5 percent). Tutoring attendance varied between the two sites, with higher attendance at Site 2 (88.3 percent) than Site 1 (79.5 percent). Additionally, Blueprint students in Site 2 attended tutoring as often as they attended school overall (88.2 percent), whereas in Site 1, Blueprint attendance was somewhat lower than school attendance (84.4 percent). In Site 1, students were occasionally pulled from their last-period tutoring session for school activities such as sports, which likely contributed to the difference between school and Blueprint attendance.

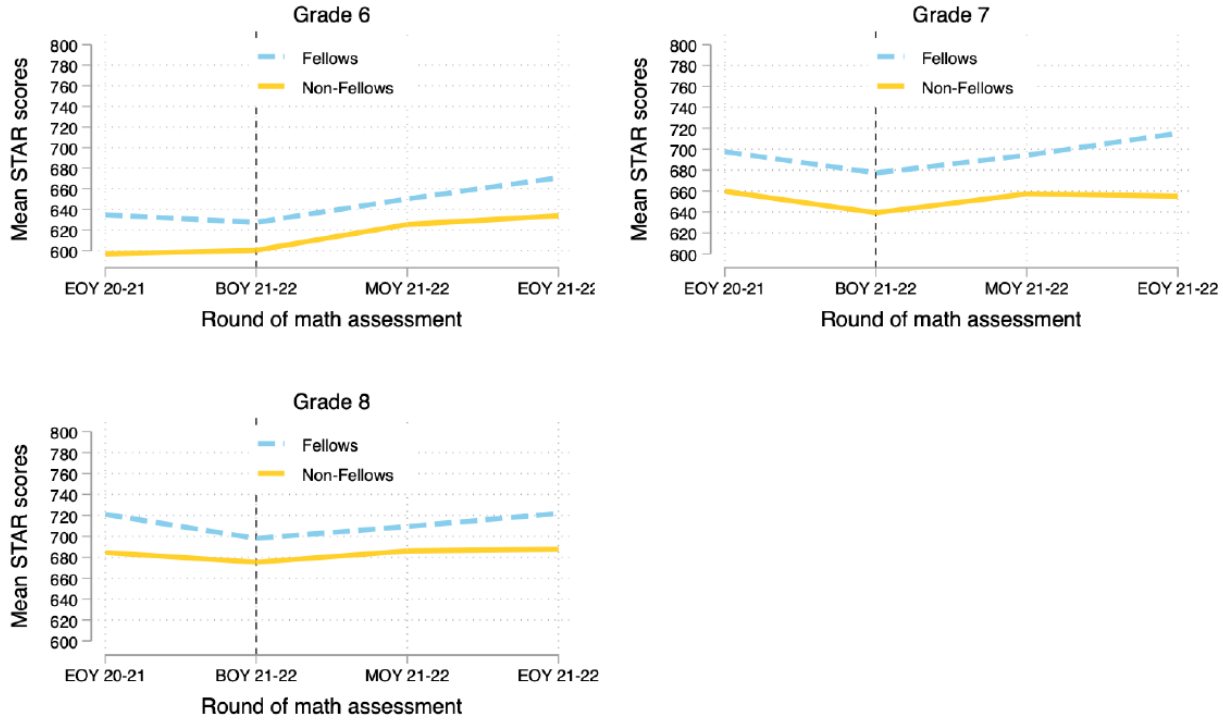
Blueprint tutoring had a moderate positive effect on student math test scores.

Blueprint participants' scores on the end-of-year Star Math assessment were 0.12 standard deviations larger than those of other students in the same schools, after adjusting for differences in baseline test scores and student grade, gender, race and ethnicity, and the presence of an individualized education program. The results are based on a Bayesian analysis of study data, and the model indicated a 99 percent probability that the program improved student scores. Despite adjusting for observable differences, it is possible unobserved differences between these groups that are related to math knowledge could affect the accuracy of our estimates. Blueprint students scored higher on the baseline assessment than comparison group students, raising the possibility of other

differences between the groups (Figure 1). See Appendix B for more details on considerations for interpreting the findings.

Unadjusted results show that Blueprint students' scores increased more than comparison students' scores during the year, widening the gap present at baseline (Figure 1). Notably, Blueprint students' scores increased in both semesters, whereas average scores among comparison students increased during the fall semester—at a rate similar to Blueprint student trajectories—but then leveled off. Scores for comparison group students in 7th grade decreased during the spring semester.

Figure 1. Unadjusted change in Star Math scores by grade among Blueprint (Fellows) and comparison (non-Fellows) students



Source: District administrative data.

Note: In the graphs, the dashed blue "Fellows" line indicates scores among Blueprint students. The solid yellow "Non-Fellows" line indicates scores among comparison group students.

BOY = beginning of year; EOY = end of year; MOY = middle of year.

Results varied by site, with a larger effect in Site 1 (0.14 standard deviations) than in Site 2 (0.09 standard deviations), and they also varied by student characteristics (Table 2). Site 1 experienced more disruption to instruction during the year due to COVID-19 than Site 2, which created differences between sites in the amount of core math instruction led by a credentialed teacher. The additional disruption to instruction that students in Site 1 experienced could have created a stronger contrast between Blueprint and non-Blueprint students, contributing to the larger effect compared to Site 2.

Table 2. Estimated effects on Star Math scores, by site and student characteristics

Sample	Estimated effect (standard deviations)	Probability of a positive effect (percentage)	Sample size
Total students	0.12	99%	954
Site 1	0.14	99%	520
Site 2	0.09	90%	434
Boys	0.13	97%	491
Girls	0.10	95%	463
Grade 6	0.09	89%	361
Grade 7	0.16	98%	281
Grade 8	0.08	85%	311
Hispanic	0.13	99%	839
White, not Hispanic	0.02	55%	73

Source: District administrative data.

Note: Sample sizes include Blueprint participants (n = 156) and comparison group students (n = 799) together, including one student with missing data on site affiliation. Sample sizes for some subgroups do not sum to the total number of students due to missing data. Due to the small number of students of other races and ethnicities, we present estimates for only Hispanic and White, not Hispanic students. See Appendix B for notes on analysis methods.

Among student subgroups examined, 7th graders experienced the largest effect, notably larger than effects among 6th or 8th graders. One potential reason could be that 7th graders at Site 1 had substitute teachers for math for the entire school year. If core math instruction they received was less effective than the instruction students at other sites received from permanent teachers, this could have caused a stronger contrast in the quality of overall math instruction (core plus tutoring) that Blueprint students received, relative to their peers.

Additionally, the estimated effect was slightly larger for boys than girls. Among Blueprint and comparison students, girls improved more than boys on the Star Math assessment during the school year. However, the gap in improvement was much narrower among Blueprint students than among comparison students. Among Blueprint students, girls' performance on the Star Math assessment improved 40.2 scale score points from the beginning to the end of the year, on average, compared to 33.8 points for boys, a 6.4-point difference. Among comparison students, girls improved an average of 29.0 points compared to 14.0 for boys, a 15.0-point difference. This suggests Blueprint tutoring may have contributed to narrowing the performance gap between boys and girls.

Reported strength of student–tutor relationships varied across sites.

Most students (76.8 percent) reported strong relationships with their tutors. We considered reported relationships to be strong when students reported an average of 4 or higher on a 5-point scale, indicating that they agreed or strongly agreed with a variety of statements suggesting a positive relationship with their tutors.

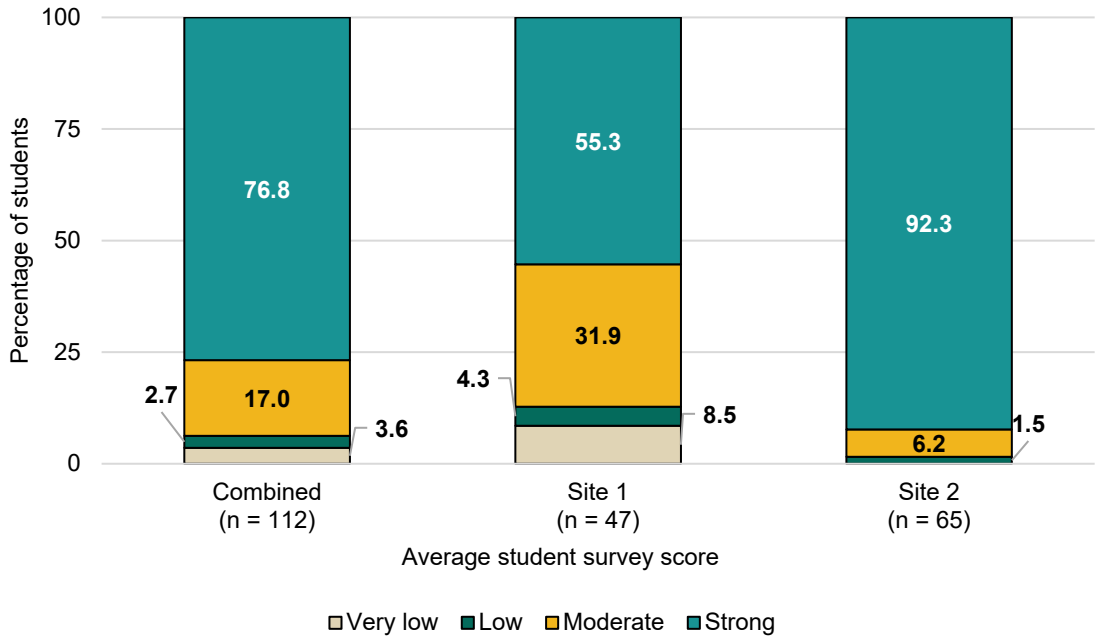
Blueprint students worked with multiple tutors during the 2021–2022 school year, which complicates the interpretation of student–tutor relationship reports. In a complement to the main study design that selected some students to receive tutoring and others not to receive it, Blueprint randomly assigned tutoring recipients into groups for homebase sessions. Students attended group sessions with their

homebase tutor twice a week, with remaining sessions split between group tutoring sessions with rotating in-person and remote tutors and independent work on ALEKS. Thus, it is not clear which tutor students referred to in their responses.

Additionally, Blueprint operated a hybrid in-person and virtual approach to group tutoring in the 2021–2022 school year. Blueprint staff noted that students seemed to enjoy in-person tutoring but felt awkward during virtual sessions where they were physically next to other students in their tutoring groups but interacting with the tutor virtually. Staff anecdotally reported it to be more difficult to develop a strong rapport with students virtually compared to in person.

Student reports of their relationships with their tutors varied substantially across sites. A much higher share of students reported strong relationships in Site 2 (92.3 percent) than in Site 1 (55.3 percent, Figure 2). In addition, 8.5 percent of Site 1 students reported very low-quality relationships, with an average score less than 2. By contrast, no students in Site 2 reported an average score less than 2.

Figure 2. Reported student–tutor relationships, overall and by site (spring 2022)



Source: Blueprint spring 2022 student survey.

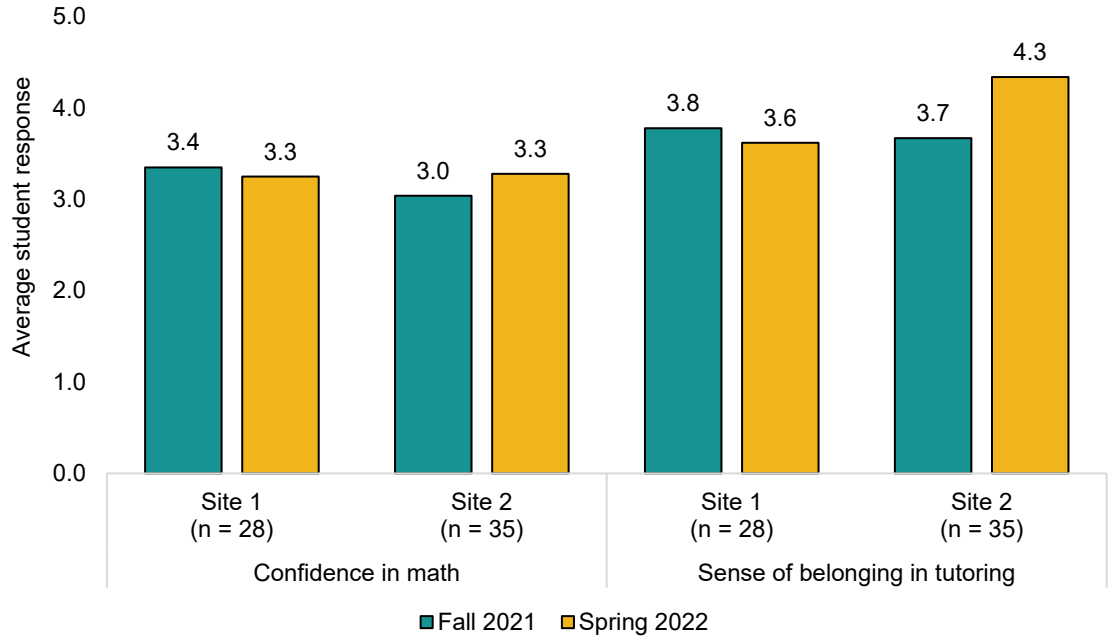
Notes: Results reflect the complete set of students who took the spring survey. The response rate was 62.2 percent overall, 56.0 percent in Site 1, and 67.7 percent in Site 2.

Several factors could help explain the different relationship results across the two sites. First, Blueprint's training is designed to ensure consistent instructional practices across tutors, and although building strong relationships is part of the training, Blueprint staff believe tutor personality is a core driver of how students perceive their tutors. They noted that the tutors at Site 1 appeared less effective at cultivating strong relationships with their students. Second, Site 1 was substantially under-enrolled at the start of the year, and many students were added to the tutoring class throughout the fall semester. Finally, Site 1 also had tutor turnover in the first two weeks of the program. The late enrollment of so many students plus the tutor turnover could have hampered the development of strong student–tutor relationships at that site.

Student confidence in math and sense of belonging in tutoring increased modestly during the school year overall but varied by site.

Among the Blueprint participants who responded to both the beginning-of-year and end-of-year surveys (approximately one-third of all participants), students reported an average increase of about 0.1 points on a 5-point scale in their confidence in math and 0.3 points in their sense of belonging in tutoring. Responses varied across sites, with reported confidence and sense of belonging in tutoring decreasing slightly for students in Site 1 and increasing for students in Site 2 (Figure 3).

Figure 3. Average student-reported confidence in math and sense of belonging in tutoring, by site (beginning and end of year)



Source: Blueprint fall and spring 2022 student surveys.

Notes: Results reflect the complete set of students who took both the fall and spring surveys. The baseline response rate was 52.3 percent overall, 50.0 percent in Site 1, and 53.8 percent in Site 2. The attrition rate was 31.5 percent overall, 20.0 percent in Site 1, and 38.6 percent in Site 2.

Compared to their peers in smaller homebase groups, students ending the year in large homebase groups exhibited the largest increase in math test scores but reported tutor relationships that were not as strong.

Descriptive analysis indicated that students ending the year in large homebase groups (eight to 10 students) showed greater improvement in Star Math scores (43.7 points on average) than either medium-sized groups (five to seven students; 34.4 points) or small groups (two to four students; 33.2 points) (Table 3). A smaller share of students in large homebase groups (70.3 percent) reported strong relationships with their tutors, compared to students in small groups (77.3 percent) or medium-sized groups (83.9 percent). By contrast, students in medium-sized homebase groups reported a meaningful improvement in sense of belonging in tutoring (0.8 response scale points, compared to no substantial change for students in other group sizes). There were no substantial differences by group size in attendance or change in reported confidence in math.

Although Blueprint randomly assigned students at the beginning of the year to either small (about four students) or large (about eight students) homebase groups, in practice, group size varied over time based on attendance and students moving out of the district. The analysis uses the size of the group students were in at the end of the year to reflect their experience of implementation. Blueprint staff reported that most changes to group sizes occurred in October, and that end-of-the-year group sizes are fairly reflective of the group sizes students experienced throughout the year. Taken together, these descriptive results suggest larger homebase groups did not appear to harm math knowledge gains, but small or medium-sized groups may better facilitate strong student–tutor relationships.

Table 3. Student outcomes by size of homebase group

Outcome	Size of homebase tutoring group			Sample size
	Small (2–4)	Medium (5–7)	Large (8–10)	
Percentage of tutoring sessions attended	84.0%	79.8%	82.5%	185
Mean unadjusted STAR Math improvement (scale score points)	33.2	34.4	43.7	156
Percentage reporting a strong relationship with tutor (spring survey)	77.3%	83.9%	70.3%	112
Change in reported sense of belonging in tutoring (on a scale of 1–5)	0.0	0.8	0.1	63
Change in reported confidence in math (on a scale of 1–5)	0.0	0.1	0.2	63

Source: District and Blueprint administrative data and student survey data.

Overarching Conclusion and Next Steps

Blueprint's tutoring program was associated with a modest positive effect on student math test scores and improvement in students' confidence in math and sense of belonging in tutoring. Effects varied across sites, potentially reflecting differences in the instructional environment between the two sites and personal differences in the tutors assigned to each site. Using a hybrid model of both in-person and remote tutors in one classroom was a significant departure from Blueprint's traditional model that yielded important lessons and is informing Blueprint's model for the 2022–2023 school year at Sites 1 and 2. Blueprint staff suspect relationships between students and tutors would be stronger in implementation models in which students remain with a consistent tutor all year.

Blueprint made several changes to its model for the 2022–2023 school year, based on lessons learned from this study. All tutoring at Sites 1 and 2 will take place in person. The program no longer uses online tutors for middle school students but will continue to use remote tutors for the high school programs. Students continue to use ALEKS as a supplement to in-person instruction, as it supports data-based instructional groupings and provides opportunities for structured independent work. Blueprint uses larger student–tutor ratios (between six and eight students per tutor) given the findings presented earlier in the report. This reduces the overall cost per student from the traditional 4:1 model and allows it to serve more students.

References

Bruch, J., Forde, J., Lai, I., & Sotelo Muñoz, V. (2022). *Menu of measures: High-quality measures of Middle Years Math student outcomes*. Mathematica. **https://www.mathematica.org/-/media/publications/pdfs/education/2022/menu-of-measures_math_march-2022.pdf**

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*. National Bureau of Economic Research. **https://www.nber.org/papers/w27476**

Renaissance Learning, Inc. (2021). *Star Assessments for Math technical manual*. **https://vdocuments.site/star-assessments-for-math-technical-manual-rp-introduction-design-of-star.html?page=1**

Appendix A. Detailed description of program

Overarching Goal of the Portfolio

The goal of the Bill & Melinda Gates Foundation's Middle Years Math portfolio is to dramatically improve middle years math instruction so that all students who are Black, Latino, and/or experiencing poverty deeply know and are able to use and enjoy math by the time they are in high school. As part of this effort, the foundation funded a study of Blueprint's Math Fellows program to develop new evidence on the program design, its implementation, and its effectiveness.

Key Components of the Program

According to the developer, the goal of Blueprint's hybrid math tutoring program is to ensure that all students receive the math instruction necessary to rise to their full academic and social potential. Blueprint aims to level the playing field so that every student, regardless of race and income status, has access to high-quality instruction and academic support. The Blueprint tutoring program does this by drawing on the strengths of in-person and online tutoring and web-based math learning platforms, and by providing incentives to support participation and family engagement.

Students attend the Blueprint Math Fellows tutoring program as part of their regular school day. Classes (tutoring sessions) are 45 minutes long and take place each school day. Blueprint tutors, called "fellows," engage with families and provide parents with regular reports and communication about student progress and goal-setting activities, supporting parents to stay engaged with their children's learning. Tutors also send text message reminders and make individual calls to families as needed to encourage student attendance and participation.

The tutoring program that Blueprint piloted in this study was an innovation that combined elements of the core in-person model used before the pandemic and elements from the remote tutoring program that Blueprint tested in the spring semester of the 2020–2021 school year when all students in the two middle schools were learning from home. Two in-person tutors and one remote tutor shared instructional responsibility for an entire class (20–25 students). The composition of student instructional groups changed regularly based on performance data from ALEKS and interim assessments. In-person tutors delivered lessons that align generally with each school's scope and sequence. The remote tutor provided online, individual support to students who are working independently on ALEKS topics during class. Although instructional groupings were fluid, students were also assigned to one of these three tutors for the entire school year for additional activities. Twice weekly, this "homebase" tutor conducted number talks and goal-setting, team-building, and motivational activities with students in one of three group sizes: large (eight to 10 students), medium (five to seven students), or small (two to four students).

Blueprint hypothesized that the key program components would, in combination, achieve short- and long-term outcomes. First, in theory, students' participation in the tutoring would foster high-quality relationships between students and tutors. Second, students' participation in tutoring and the support students receive from tutors would increase students' sense of belonging and confidence in math and would encourage students to feel comfortable asking questions and taking risks in the tutoring and core classroom environments. Third, students' participation in the tutoring and their effective use of the ALEKS platform would lead to improvements in students' math knowledge and improved grades in math courses. Short-term improvements in student perceptions of math and math performance were expected to translate into long-term impacts, including an increase in students'

participation in regular math classes, an increase in students' willingness to challenge themselves and take advanced math courses, and improved math performance in subsequent academic years.

Both middle schools in the district participated in the study. Participating students were in grades 6, 7, and 8. The program lasted the entire school year (38 weeks in total), and 177 students participated in the program. Most of the students who left the program early did so because their family moved out of the district.

Appendix B. Methods

Quasi-experimental design. School administrators at two sites in a suburban school district assigned students in grades 6, 7, and 8 to Blueprint tutoring, primarily based on student schedules. Students in the same schools and grades who were not assigned to Blueprint tutoring received other elective or intervention instruction and formed the study's comparison group. Blueprint tutoring did not replace core math instruction in students' schedules. Students participated in the Blueprint for the entire school year. When program vacancies appeared from students transferring out of the school, school administrators assigned additional students to the program. At the beginning of the school year, Blueprint randomly assigned students to large homebase groups (eight students) or small homebase groups (four students). During the year, group sizes that students experienced changed, as students moved into and out of the program.

Measures. Blueprint staff collected student survey data at the beginning (fall) and end (spring) of the school year from students receiving tutoring. Blueprint used the Copilot-Elevate tutor caring subscale (student–tutor relationship construct), Copilot-Elevate classroom belonging subscale (sense of belonging construct), Patterns of Adaptive Learning scale (PALS) (math confidence construct), and Math and Me (math confidence construct) to measure student perceptions about the tutoring and their experience of math. Details on each of the survey measures are available in Mathematica's menu of high-quality middle years math student outcome measures, which were selected in consultation with external measurement experts (Bruch et al., 2022). For pre-post analyses of student math confidence and sense of belonging in tutoring, we used data from students who completed both surveys. For post-only descriptive analyses of student–tutor relationships, we used the full set of spring survey data and thus drew on a larger sample than the pre-post analyses. The fall student survey included data from 92 students, a 52.3 percent response rate. The spring survey included data from 112 students, a 62.2 percent response rate.

The school district administered the Star Math assessment to Blueprint and comparison group students at the beginning and end of the school year and provided the data to the study team. The response rate in the baseline Star Math administration was 87.2 percent and was similar between students participating in Blueprint (85.9 percent) and those in the comparison group (87.4 percent).

Analysis. We used descriptive techniques to analyze reported student–tutor relationship strength, calculating the percentage of students with spring survey data who reported an average of at least 4 out of 5 when asked their level of agreement with statements describing elements of positive student–tutor relationships. To analyze pre-post student survey measures (reported math confidence and sense of belonging in tutoring), we calculated mean reported scores on scales of 1 to 5 for students with both fall and spring data and subtracted the average fall score from the average spring score to show change over time.

The study team conducted regression-adjusted comparison group design analyses, implemented by an analyst at the empirical analysis group RISC, to estimate the effects of Blueprint tutoring on student test scores. We applied Bayesian shrinkage to the resulting estimates, assuming a zero-centered prior distribution with a standard deviation based on a summary of effect sizes from recent, rigorous education program evaluations. This enabled us to produce Bayesian estimates of the effect size and associated probability of a true positive impact.

As a sensitivity analysis, for each grade, we calculated matched comparison group design Bayesian estimates—which incorporate information from evidence of effectiveness from prior studies of education interventions. We then produced a weighted average of the effect across grades to compare with our main report finding on math knowledge. The weighted average is similar to the main finding (0.08 compared to 0.12 standard deviations, Table B.1), bolstering confidence in our main finding estimate. The differences between these estimates are likely due to two causes. First, the main finding that was not based on matching might reflect some inaccuracy due to the analysis sample including students whose characteristics were too dissimilar from tutoring recipients to be included in the matched sample. Second, the weighted average from the matched comparison group estimates is likely a slight underestimate due to reduced precision resulting from smaller sample sizes in the grade-specific estimates we used to derive the weighted average. This reduced precision causes the Bayesian analysis to give relatively more weight to prior evidence because this weight is larger when the matched comparison estimate is less precise and is therefore adding less new information.

Table B.1. Grade-specific and weighted average matched comparison group design Bayesian estimate of the effect of Blueprint tutoring on Star Math scores

Grade	Estimate (standard deviations)	Blueprint sample	Comparison group sample	Total sample
6	0.00	50	44	94
7	0.12	51	46	97
8	0.12	55	43	98
Weighted average	0.08	156	133	289

Source: District administrative data.

Limitations. Attrition from the baseline to the end-of-year surveys was 31.5 percent and was greater in Site 2 (38.6 percent) than in Site 1 (20.0 percent). However, baseline differences in math confidence and sense of belonging measures between students with and without follow-up survey data were greater in Site 1 (Table B.2).

Table B.2. Student survey baseline differences, by site

	Site 1 (Baseline n = 35)			Site 2 (Baseline n = 57)		
	With follow-up data (n = 28)	Without follow-up data (n = 7)	Difference	With follow-up data (n = 35)	Without follow-up data (n = 22)	Difference
Math confidence	3.2	3.8	-0.6	3.1	3.0	0.2
Sense of belonging in tutoring	4.0	3.1	0.9	3.7	3.7	0.0
Sense of belonging in math class	3.8	3.1	0.7	3.7	3.7	0.0

Source: Fall student survey.

The most important potential source of inaccuracy in the analysis of math knowledge arises from the possibility that students who participated in Blueprint were different from those who did not in ways

that affected their growth in math knowledge during the year. Specifically, although district administrators selected students for Blueprint based on their schedules and not on observed differences in math performance, Blueprint students scored higher on the baseline assessment than comparison group students (Figure 1 in the main report). The difference in baseline scores was substantial—more than 20 test score points in each grade. This magnitude is similar to the overall average change in scores over the course of the school year. It is possible Blueprint students' math performance would have improved at a faster rate than that of comparison group students even in the absence of tutoring, and statistical adjustments based on baseline test scores and student characteristics may not fully address this source of potential inaccuracy. The similarity between regression-based estimates and matched comparison estimates provides some evidence that differences in baseline math knowledge and other characteristics in the regression sample did not play a large role in differences in math knowledge in spring 2022. However, there still could have been differences in the characteristics of Blueprint and comparison students that the study did not observe that nonetheless explain some of the differences in spring 2022 math knowledge but that our estimates attribute to Blueprint tutoring.