AIR®
Advancing Evidence.
Improving Lives.

Results Summary

Photo by Allison Shelley for EDUimages

Balance the Equation Algebra I Grand Challenge Findings

Toni Smith | Martyna Citkowicz | Cassandra Meyer

January 2026

The Gates Foundation launched the Balance the Equation Algebra I Grand Challenge (AGC) in fall 2020 to identify and study innovative solutions aimed at improving math outcomes for students who are Black, Latino, affected by poverty, or designated as English learners (priority students). Student outcomes of interest included positive experience in math class, positive math identity, and math proficiency. The Foundation funded 15 grantees in spring 2021 to refine and design a study of their solution for improving student outcomes and funded 10 of those grantees in summer 2021 to conduct the studies of their solutions.[1] The American Institutes for Research® (AIR®) was a learning partner to the Foundation and the AGC grantees, supporting grantees in study planning and execution. Each study was designed to answer questions about solution implementation, effect on student outcomes, and scalability. AIR supported data collection and analysis activities. Three grantees implemented their solutions and collected data during the 2021–22 school year (Cohort 1), six grantees during the 2022–23 school year (Cohort 2), and one grantee during the 2023–24 school year (Cohort 3). This executive summary provides an overview of the solutions; the research questions, study designs, and data sources; the challenges faced; and the summary findings from the studies.

Solution Descriptions

Grantee solutions included teacher professional development programs, student programs delivered by grantee-trained teachers and/or mentors, and student programs delivered by grantee staff or a technology platform. Exhibit 1 shows the grantees in each cohort, the type of solution for each grantee, and a description of each grantee's solution. Three solutions focused on teacher professional development, five solutions on a student program delivered by a trained teacher and/or mentor, one solution on a student program delivered by grantee staff, and one solution on a student program delivered by a technology platform.

[1] The Foundation funded one additional grantee to study their solution. The grantee experienced difficulty recruiting a district partner and, thus, did not complete their study.

Exhibit 1. Grantees, Solution Type, and Solution Description for Each Cohort

Cohort	Grantee	Solution type	Solution description
Cohort 1	Better Lesson (BL)	Teacher professional development	Teacher professional development designed to support use of new instructional strategies and curricular adaptations to meet student needs in high school Algebra I.
Cohort 1	Howard University Public Charter School of Math and Science (Howard)	Student program delivered by trained teachers	A series of Illustrative Math lessons that were revised to be learner centered for eighth-grade Black girls. Teachers are trained to deliver the lessons.
Cohort 1	Village Education Tutors Foundation (VETF)	Student intervention delivered by grantee staff	After-school tutoring sessions focused on math and wellness for eighth-grade Black and Latina girls.
Cohort 2	Black Teacher Collaborative	Teacher professional development	Professional development designed to support use of new instructional strategies and curricular adaptations to improve academic outcomes of Black students learning Algebra I, when taught by Black teachers.
Cohort 2	ConnectED	Teacher professional development	Professional development focused on Math Language Routines (MLRs) and delivered through a Video Club where teachers collaboratively analyze Algebra I class videos with a tagging tool to highlight teachers' and students' use of MLRs, followed by discussions and reflections about the observed teaching strategies.
Cohort 2	The Curtis Center at the University of California Los Angeles (Curtis Center)	Student program delivered by trained teachers and mentors	Semester-long supplemental math program for seventh-, eighth-, and ninth-grade students in which students carry out a university-like research investigation of a currently relevant and authentic applied mathematics question, with advisory by a varied group of mathematicians as well as University of California Los Angeles undergraduate science, technology, engineering, and math majors/mentors. Teachers are trained to deliver the program.
Cohort 2	Rhode Island Department of Education (RIDE)	Student program delivered by trained teachers	A semester-long supplemental course for eighth-grade math students that provides 4 additional hours of in-school math instruction each week to reinforce key concepts necessary for success in Algebra I through student-centered discussions, instructional approaches to support English learners, and the development of productive persistence skills. The course was developed by an external group and teachers are trained to deliver the course.

Cohort	Grantee	Solution type	Solution description
Cohort 2	Young People's Project (YPP)	Student program delivered by trained mentors, with teacher support	High school students (Math Literacy Workers) are trained to facilitate near-peer math workshops with middle school students using a game-based curriculum. Teachers are trained to monitor student activity.
Cohort 2	Zearn	Student program delivered by grantee technology platform	Use of videos that align to student experience for remediation in the Zearn digital math curriculum.
Cohort 3	University of Florida (UF)	Student program delivered by trained teachers	A math program that replaces two units of the Illustrative Mathematics Algebra I curriculum and uses the Math Nation online platform to provide (a) algebra content aligned to student experience, (b) opportunities for students to network together, and (c) mechanisms for students to track their math learning. Teachers are trained to deliver the program.

Research Questions

Exhibit 2 shows the research questions (RQs) for the AGC studies. Implementation RQs focused on the extent to which the solution was implemented as planned during the study and participant perspectives of the solution. Outcomes RQs focused on the effect on student outcomes and, if applicable, teacher and/or mentor outcomes. Scaling considerations RQs focused on the conditions for implementation, solution revisions, and solution cost.

Exhibit 2. AGC Study RQs

Research questions
Implementation 1. To what extent are the solution activities implemented as intended (e.g., frequency and coverage)? 2. To what extent do participants complete the activities? 3. What are participants' perspectives on the usefulness of solution activities and the usability of solution materials?
Outcomes 4. (If applicable) Do trained teacher/mentor beliefs about student math learning, use of effective math instructional practices, and skill in adapting curricula and instruction to meet student needs change from before to after their participation in the solution? 5. Does participation in the solution relate to/have an impact on students' experiences in math class, math identity, and math proficiency?
Scaling considerations 6. What are the necessary conditions for successfully implementing the solution? 7. What does the grantee recommend for an updated/revised solution, and what would it cost to implement it?

Grantees identified and implemented their solutions for the studies with one or more school districts located in different regions in the United States. AIR supported grantees to secure research approvals and the necessary consent for data collection.

Study Designs and Data Sources

We used a mixed-methods study design to address RQs 1–3. We collected grantee records documenting solution activity and attendance, teacher logs documenting solution implementation, platform backend data documenting student activity in a technology platform, and teacher and student focus groups and/or surveys, as applicable. We conducted descriptive analyses of the solution records, teacher logs, and platform backend data. We audio recorded and transcribed the recordings of the focus groups. Then, we conducted qualitative analyses of those data.

When applicable, we used a pre–post study design to address RQ 4. We administered a teacher/mentor survey on beliefs about student math learning, use of effective math instructional practices, and skill in adapting curricula and instruction before and after experience with the solution. We computed the average construct score across survey respondents on each outcome for each survey administration to examine changes in outcomes from before to after solution participation.

For each grantee, we used either a correlation or comparison study design to address RQ 5. Six of the 10 grantees used a correlation study design and four used a comparison study design.

To collect data on aspects of student math experience and identity, we used a student survey. To collect data on math proficiency, we collected from the participating district(s) student math scores on formative and end-of-year state assessments, as well as course grades. We also collected from districts student demographic data.

For correlational study designs, we calculated solution dosage—defined as time spent in and/or number of solution activities completed/attended—for each student.[2] For teacher professional development programs, we defined student dosage as the number of hours for which their teacher participated in solution activities. Then, we estimated the relationship between dosage and aspects of math class experience, identity, and proficiency using an analytic model that controlled for demographic characteristics and the baseline measure of the relevant outcome.[3,4]

[2] In some cases, we gave more weight to specific solution activities based on consultation with the grantee regarding activity importance.
[3] We used data from the student survey administered prior to experience with the solution to construct the baseline measure for analyses of effect on survey outcomes and prior year state standardized math test scores as the baseline measure for analyses of the effect on math proficiency outcomes.
[4] In two of the six correlational studies, most students were missing prior year math proficiency data; thus, we were unable to control for baseline math proficiency in analyses of math proficiency outcomes in these two studies.

For comparison study designs, we estimated the difference in each student outcome between participants in the solution (treatment students) and nonparticipants (comparison students) using an analytic model that controlled for demographic characteristics and a baseline measure.[5,6]

After conducting the individual grantee analyses, we pooled findings on student survey and math proficiency outcomes across grantees using meta-analysis.[7,8] Meta-analysis allows for increased statistical power to detect significant effects by combining results from multiple studies, thereby leveraging the larger, combined sample size. We conducted three sets of meta-analyses.[9] The first set pooled the overall findings, with grantees grouped by solution type and study design: teacher professional development solutions that used a correlation study design, student-facing solutions that used a correlation study design, and student-facing solutions that used a comparison study design. The second set pooled the effects for the priority students who were the focus of the AGC, with grantees grouped by these same three study design categories.[10]

The third set of meta-analyses utilized solution emphasis coding to identify specific instructional or structural elements that may relate to stronger student outcomes. Specifically, we examined pooled effects for solutions that addressed each of the focus areas identified by the AGC solicitation, which included strengthening teacher practice, elevating understandings of math language, building out student support systems, developing new or better feedback mechanisms, and improving the relevance of Algebra content for students. We also pooled results for solutions that addressed math content prior to Algebra I and those that addressed math content associated with Algebra I. To gain insight into which solution emphases relate to student outcomes, we conducted these exploratory meta-analyses on the student outcomes for which the group of solutions showed positive or null impacts or relationships.[11] This exploratory approach ensures the synthesis highlights solution features that demonstrate a promising and consistent signal of benefit, focusing the analysis on instructional elements that show a potential for improvement without evidence of substantive adverse effects.

Finally, we conducted interviews with district leaders to address RQ 6 and with grantees to address RQ 7.

---

[5] Three of the four comparison studies included a quasi-experimental design in which we used a statistical matching algorithm to select comparison students similar to treatment students. The fourth comparison study randomly assigned students to treatment and comparison groups; thus, no matching was needed. For all four studies, we assessed baseline equivalence to ensure treatment and comparison students were similar at baseline on baseline pretest measures and demographics.

[6] For analyses of effect on student survey outcomes, for two of the four comparison studies, the baseline pretest measure was created from the baseline student survey responses. For the other two comparison studies, comparison students did not complete a survey prior to experiencing the solution; thus, the baseline pretest measure on student survey outcomes was the prior year state standardized math test scores. For analyses of effect on math proficiency outcomes, the baseline pretest measure was the prior year state standardized math test score for all four comparison studies.

[7] Meta-analysis is a statistical technique that combines results from multiple effects (or grantees) by weighting the contribution of each estimate of the effect based on the statistical precision with which that effect was estimated. Effects that are estimated from a more precise sample (generally, the study sample with the largest number of participating students) are weighted more heavily in the pooled effect, allowing the larger sample to contribute more information in calculating the overall pooled effect.

[8] To combine the effects of each outcome across grantees, we standardized all estimates by dividing each estimate by the sample standard deviation of the outcome in the raw metric. We then converted the standardized estimates into the Pearson correlation metric by multiplying each estimate by the standard deviation of the treatment indicator—for correlational studies, this was the standard deviation of the dosage measure, and for comparison studies, we assumed a standard deviation of 0.50.

[9] We ran all three meta-analyses using a fixed-effects model, which uses the inverse of the within-grantee variance to weight the estimates.

[10] For each grantee, we estimated effects specifically for the priority students targeted by the AGC: Black students, Latino students, English learners, and students experiencing poverty. Students only needed to belong to one of the priority groups to be included in the priority group analyses. The underlying individual grantee analyses for this group used the same models as those used for the overall full sample analyses, with the data subset solely to priority group students.

[11] We only pooled effects for a specific outcome-feature combination if the observed effects for the contributing grantees met the following criteria: the set had to include at least one effect greater than 0.10 and could include null effects (between -0.10 and 0.10), but it could not include any negative effects of -0.10 or below.

Challenges Faced

Each of the studies encountered challenges related to recruitment, data collection, and/or implementation. Solution implementation occurred during the 2021–22, (Cohort 1), 2022–23 (Cohort 2), and 2023–24 (Cohort 3) school years, as districts and schools were experiencing the aftermath of the COVID-19 pandemic as well as continued teacher shortages. This made buy-in to new initiatives and study activities difficult to secure, particularly for Cohorts 1 and 2. In addition, many districts required active parent/guardian consent for participation in study student surveys and focus groups. As a result, eight of the 10 studies had fewer participants than recommended for the study design. Further, survey response rates were low, even among consented participants, and district demographic and achievement data were limited for reasons that varied among districts. The analytic samples for each study were even smaller, as a result. Finally, implementation was delayed in some cases due to recruitment difficulties and other disruptions. These challenges should be considered when interpreting the results.

Implementation Findings

Solution implementation varied across studies, with some studies having close to full implementation of solution activities and others having weaker implementation. Regardless of the level of implementation, solution participants generally had positive impressions of the solutions and their potential for having an impact on student math outcomes.

For four of the 10 studies, solutions were generally well implemented, with almost all of the teacher and mentor training delivered and students experiencing all, or almost all, of the student-facing activities, as applicable. Implementation for the other six studies was not as strong, due to challenges with teacher buy-in and burden, difficulty scheduling solution activities, and/or other district disruptions.

Participants generally enjoyed their experience and felt that the solution activities would support student math learning. Teachers in the three of the studies noted that it would have been helpful to have more support for translating learning from the training sessions to use with students.

Teacher and Mentor Outcomes Findings

Across the grantees who used correlational study designs, 43 teachers completed surveys both before and after solution participation.[12] For comparison study designs, 19 teachers and 35 mentors completed surveys at both time points.[13]

For the correlation studies, teachers reported making adaptations to meet student needs and using effective practices more frequently after participating in the solution in three of the studies but less frequently or similarly in two of the studies.

While teacher and/or mentor beliefs generally stayed the same across comparison studies, we identified improvements in three specific areas across two studies. One study showed improvements in both beliefs about student math learning and the use of effective instructional practices. A separate study showed improvements in the use of adaptations.

[12] One of the solutions using a correlational study design did not include teacher participants; thus, the teacher survey findings are based on five of the six correlational studies.
[13] Only two solutions included mentors; thus, the mentor survey findings are based on two of the four comparison studies.

Student Outcomes Findings

Across the grantees, the student outcomes sample for teacher professional development solutions included 74 to 3,037 students. For student-facing solutions, the sample size ranged from 187 to 615 students for correlational study designs and 477 to 1,076 students for comparison study designs.

The effects on student math learning were mixed for the correlational studies. The relationship between dosage for solution activities and course grades was not statistically significant and close to zero for three solutions, positive but not statistically significant for one solution, and statistically significant and negative for one solution. The relationship between dosage and state assessment scores was also not statistically significant for three solutions but statistically significant and positive for two solutions. For two solutions, the relationship with formative assessment scores was statistically significant and positive. The relationships between dosage and student outcomes related to math class experience and math identity were generally positive and statistically significant for one solution, close to zero for one solution, and negative but generally not statistically significant for the other three solutions. Due to challenges collecting data for one solution, we could not estimate the relationship between dosage and math class experience, math identity, and other measures of math proficiency.

For the comparison studies, the effects on aspects of student math class experience and math identity were mixed but generally not statistically significant, and effects on math proficiency were generally positive or close to zero. Two solutions showed generally positive effects on math classroom experience and math identity, while the effects for two other solutions were generally negative. The studies of two solutions show positive and statistically significant effects on course grades.[14]

Meta-Analysis: Solution Type and Study Design

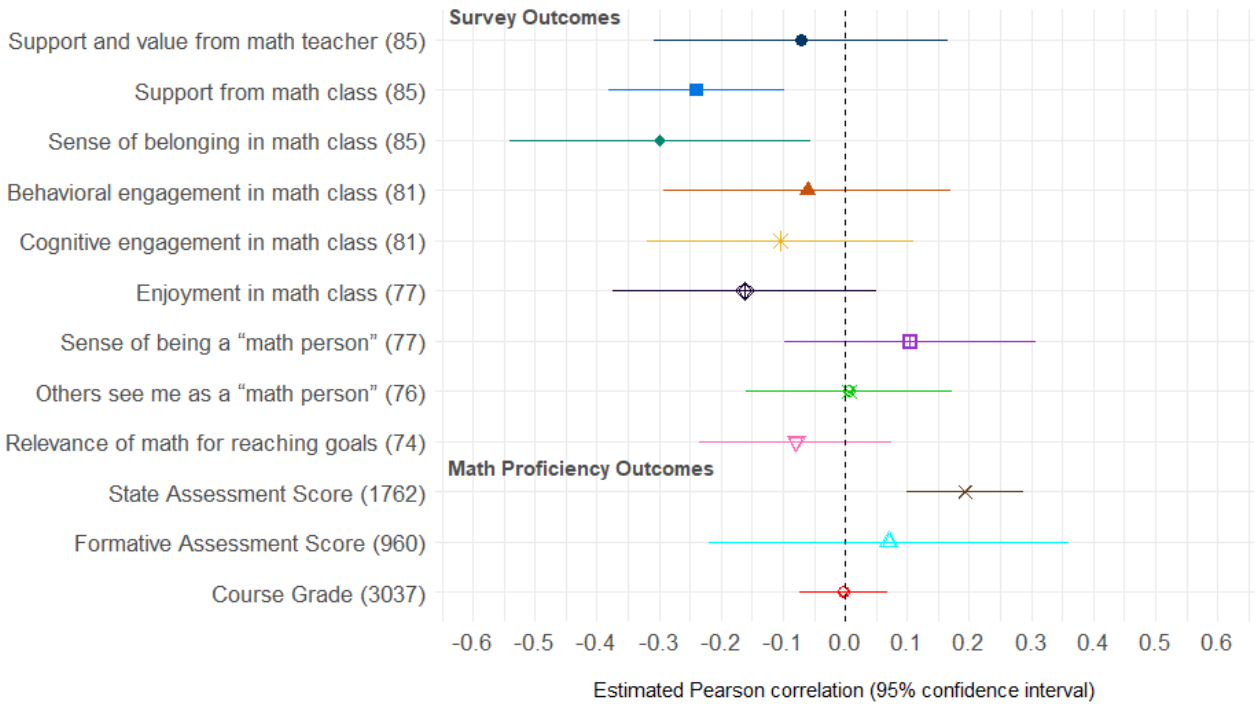
In this section, we present the pooled findings across grantees by solution type and study design. All reported effects are presented in the Pearson *r* metric, which represents the strength and direction of the relationship between the solution and the outcome. A value closer to +1 indicates a strong positive relationship, a value closer to 1 indicates a strong negative relationship, and a value closer to 0 indicates a weak or no relationship.

Exhibit 3 presents the pooled average effects of the teacher professional development solutions using correlation study design on student outcomes. We found a statistically significant positive effect on state assessment scores ($r = 0.19$), indicating a positive association between teachers' solution participation and students' math scores on the state assessment. However, the effects varied across grantees, with one showing a positive effect and one showing a negative effect.

We also found statistically significant negative effects for two survey outcomes, including students' perceptions of their support from math class ($r = -0.24$) and their sense of belonging in math class ($r = -0.30$). For both survey outcomes, the effects were consistently negative for both grantees.

[14] The estimates represent the average difference in course grade outcomes between treatment and comparison students on a 5-point grading scale.

Exhibit 3. Average Effects of the Teacher Professional Development Solutions Using a Correlation Study Design



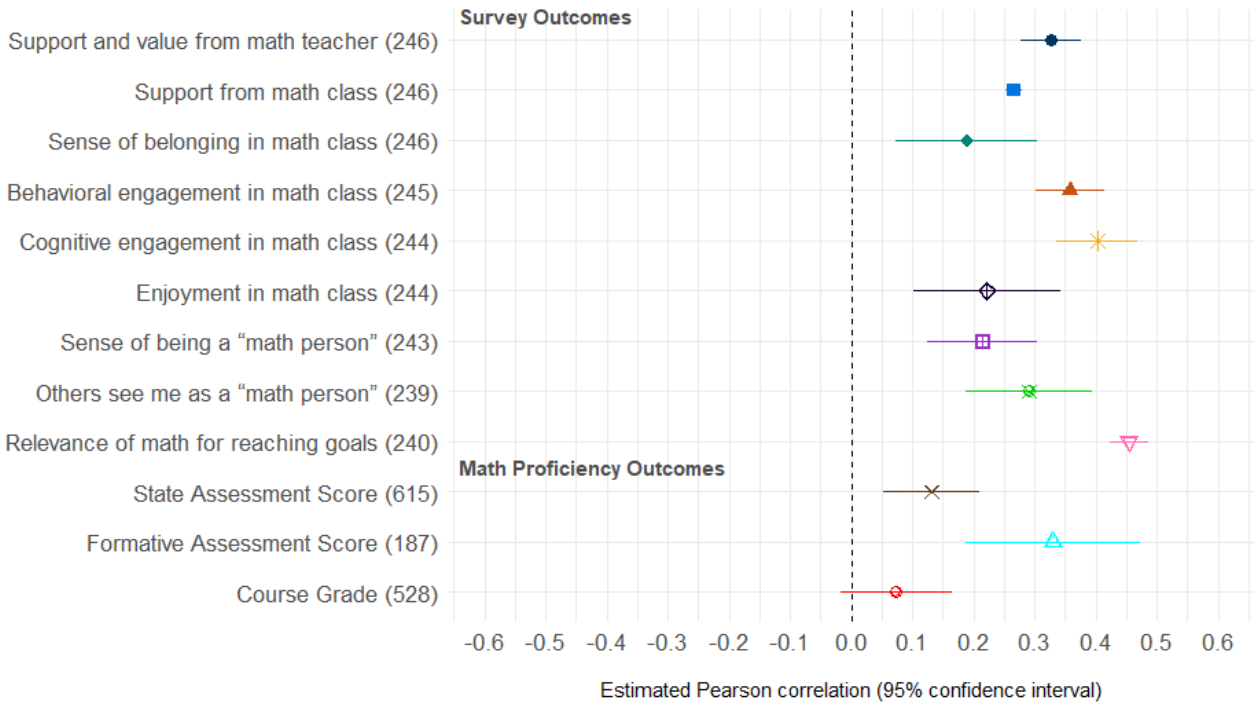
Note. Exhibit 3 presents the average effects of the teacher professional development solutions for each outcome, with total student sample size included in parentheses. The points represent the effects, and the lines represent 95% confidence intervals. Confidence intervals that include zero indicate nonsignificant effects. The number of grantees included for the specific outcome analysis varied: two grantees contributed to student surveys and state assessments, three to course grades, and one to formative assessments.

Exhibit 4 presents the pooled average effects of the student-facing solutions using a correlational study design on student outcomes. We found statistically significant positive effects on all nine survey outcomes and two of the three math proficiency outcomes, indicating a positive association between solution participation and student outcomes.

Effects on students' survey outcomes related to their math class experience and identity ranged from 0.19 to 0.45. While the pooled average effect was positive for all survey outcomes, the effects showed some variation across the grantees. One grantee had positive effects on all nine survey outcomes, the second grantee had positive effects on six of the survey outcomes, and the third grantee had positive effects on only two of the survey outcomes.

We found statistically significant positive effects on state (r = 0.13) and formative (r = 0.33) assessment scores. For both math proficiency outcomes, all effects were consistently positive for all grantees.

Exhibit 4. Average Effects of the Student-Facing Solutions Using a Correlational Study Design

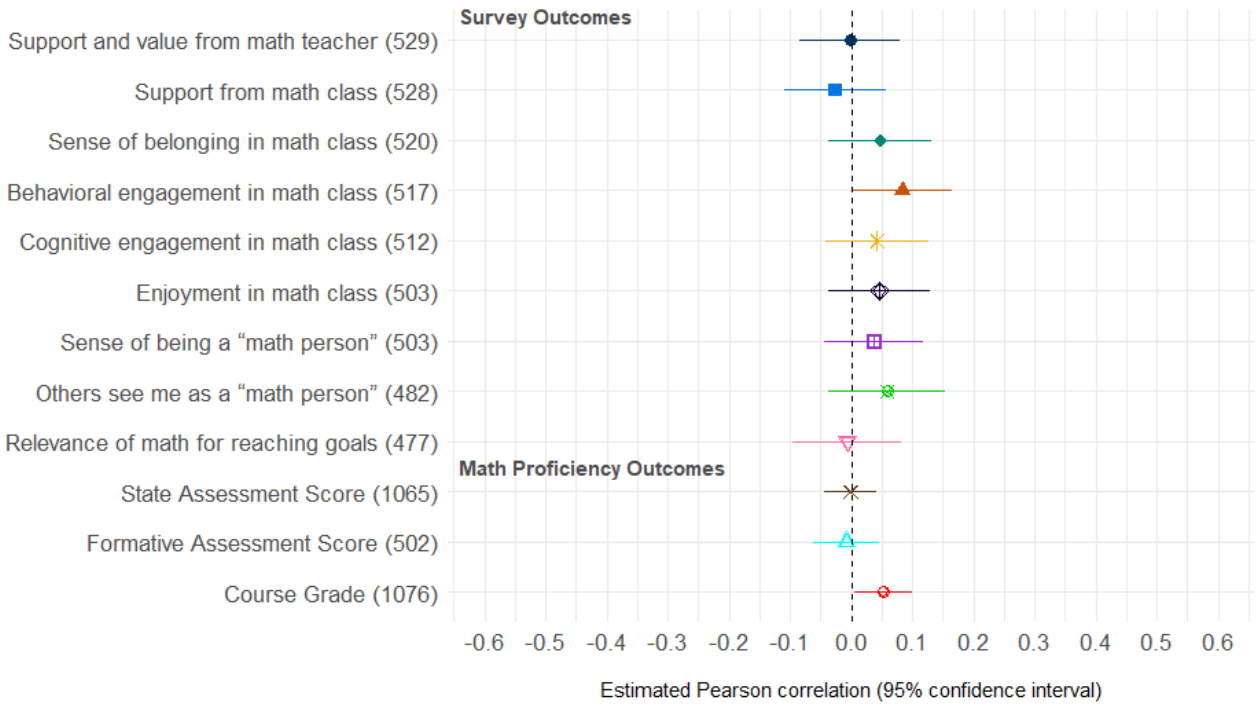


Note. Exhibit 4 presents the average effects of the student-facing solutions using a correlational study design for each outcome, with total student sample size included in parentheses. The points represent the effects, and the lines represent 95% confidence intervals. Confidence intervals that include zero indicate nonsignificant effects. The number of grantees included for the specific outcome analysis varied: three grantees contributed to student surveys and state assessments, and two different sets of grantees contributed to formative assessments and course grades.

Exhibit 5 presents the pooled average effects of the student-facing solutions using a comparison study design on student outcomes. We found two statistically significant positive effects: one on course grades (r = 0.05) and one on students' perceptions of their behavioral engagement in math class (r = 0.08). Both findings indicate a positive association between solution participation and these two student outcomes.

However, the effects for both outcomes varied across the grantees. For behavioral engagement in math class, the effects were negative for two grantees and positive for two grantees. For course grades, the effects were positive for three grantees and negative for one grantee.

Exhibit 5. Average Effects of the Student-Facing Solutions Using a Comparison Study Design



Note. Exhibit 5 presents the average effects of the student-facing solutions using a comparison study design for each outcome, with total student sample size included in parentheses. The points represent the effects, and the lines represent 95% confidence intervals. Confidence intervals that include zero indicate nonsignificant effects. The number of grantees included for the specific outcome analysis varied: four grantees contributed to student surveys, five to state assessments and course grades, and two to formative assessments.

Meta-Analysis: AGC Priority Student Groups

Across grantees, students in the AGC priority group comprised the majority of the sample—and in some cases, the entire sample.[15,16] As a result, the pooled findings for the priority group sample generally mirror those for the full sample. For teacher professional development solutions using a correlation design, we found a statistically significant positive effect on state assessment scores (r = 0.19) and a significantly negative effect on students' perceptions of their sense of belonging in math class (r = -0.29) but no significant effect on students' perceptions of their support from math class. For student-facing solutions using a correlational study design, we again found statistically significant positive effects on all nine survey outcomes (r = 0.18 to 0.46) and on two of three math proficiency outcomes, specifically state (r = 0.11) and formative (r = 0.33) assessment scores. For student-facing solutions using a comparison study design, no significant effects were found across any outcomes, including course grades or students' perceptions of behavioral engagement in math class.

[15] Priority students included Black students, Latino students, English learners, and students experiencing poverty.
[16] Sample sizes for priority students ranged from 61 to 2,462 for teacher professional development solutions, from 187 to 503 for student-facing solutions using a correlational study design, and from 461 to 1,019 for student-facing solutions using a comparison study design. They represent 86% of the full study samples, on average.

Meta-Analysis: Solution Emphasis

Our pooled analyses of AGC grantees by solution emphasis revealed varied findings. Because these exploratory analyses were intended to identify promising practices, we only pooled findings for feature-outcome combinations that demonstrated a consistent pattern of positive or null results, as outlined above.[17] While the majority of solution features demonstrated positive effects primarily within the math proficiency category, the small subset of solutions focused on elevating understandings of math language (two grantees) were unique in meeting the pooling criteria across all three outcome areas, including math class experience, math identity, and math proficiency. In contrast, while features like empowering teacher practices and focusing on Algebra 1 content were more common across the portfolio, their reliable positive effects were concentrated in math proficiency and did not meet the thresholds for pooling regarding student survey outcomes.[18]

Solutions focused on elevating understandings of math language statistically significantly improved students' perceived sense of belonging (r = 0.11), behavioral engagement (r = 0.15), and enjoyment (r = 0.11) in math class, along with their course grades (r = 0.06). Regarding math proficiency outcomes, solutions focused on Algebra 1 content had substantial and significant effects on both state (r = 0.17) and formative (r = 0.33) assessment scores, while those that build systems for support only showed a significantly large effect on formative assessment scores (r = 0.33). Small but significant improvements in course grades were observed for algebra readiness content solutions (r = 0.08) and those that empower and strengthen teacher practices (r = 0.05), with the latter also yielding a small significant effect on state assessment scores (r = 0.04). No significant effects were found for solutions focused on improving relevance of algebra content.

Scaling Conditions Findings

With respect to conditions for successful implementation, district leaders commonly noted the importance of garnering teacher buy-in, fostering teacher support and training, ensuring alignment between the solution activities and district vision and/or curriculum, and maintaining effective communication between solution leaders and all levels of district staff (e.g., district leaders, school leaders, and teachers). In some cases, interviewees also discussed the need for the solution leaders and district staff to work together on determining how best to implement the solution, such as incorporating the solution into the school schedule and assessing student needs prior to the intervention.

With respect to solution revisions, two grantees do not plan to make revisions; two grantees plan to extend the solution to 2 years; and five grantees continue to revise their solution to (a) provide more or less time for teacher/mentor support and/or (b) include expanded curricular activities. One grantee will no longer use the solution in schools.

[17] Only eight of the 10 grantees were included in these exploratory analyses. The two remaining grantees were excluded because of their poor implementation, which makes it difficult to reliably attribute observed student outcomes to the solution.
[18] The number of grantees included in each meta-analysis varied by outcome category. For math proficiency outcomes, we pooled findings for solutions that focused on empowering teacher practices (seven grantees), elevating understandings of math language (two grantees), building systems for support (four grantees), improving relevance of algebra content (four grantees), algebra readiness content (three grantees), and Algebra 1 content (four grantees). For math class experience and math identity survey outcomes, only the solutions focused on elevating math understandings of math language (two grantees) met the criteria for pooling effects. The feature on developing new or better feedback mechanisms did not meet the criteria for pooling effects in any outcome category.

Summary and Future Exploration

In summary, the 10 studies generally found that solution participants enjoyed the activities and expected them to have positive effects on student math outcomes. Analyses of the effects of individual solutions on teacher and student outcomes revealed mixed results. However, challenges with study implementation resulted in smaller sample sizes than desired. Pooled analyses offer the opportunity to use a larger sample and to examine the effects of solutions with common features. Teacher professional development solutions using a correlation study design demonstrated a positive, statistically significant effect on state assessment scores but were associated with negative, statistically significant effects on students' self-reported support and sense of belonging in math class. Student-facing solutions studied using a correlational design showed positive, statistically significant associations across all student survey outcomes related to math class experience and math identity and two of the three math proficiency outcomes, including state and formative assessment scores. Student-facing solutions studied using a comparison design exhibited small, positive, statistically significant effects on students' course grades and perceptions of behavioral engagement in math class.

Despite some mixed grantee findings, the pooled findings show promise for the solutions, particularly their potential to improve student math learning. These benefits were similarly observed for the AGC priority student group sample, which comprised the majority of the study population. Our solution emphasis meta-analysis further suggests that solutions focused on elevating understandings of math language showed the broadest promise across student survey and math proficiency outcomes, while Algebra 1 content and building systems for support emerged as strong drivers of improving student assessment scores. More work is needed to understand the broader findings on math class experience and math identity. It may be, for example, that students answering questions about math class experience were more critical of their "regular" math class given the experiences they were having in the solution activities. In addition, it would be useful to conduct similar studies now, post-pandemic, to provide more robust information about the effectiveness of these solutions during a period of fewer implementation disruptions.