The Effects of Adopting Eureka Math or Into Math on Student Achievement

A Quasi-Experimental Study of Middle School Math Curriculum Use

August 11, 2023; Updated May 17, 2024

Naihobe Gonzalez, Silvia Robles, Chloe Shawah, and Amy Johnson

Submitted to:

Bill & Melinda Gates Foundation
Attention: Andy Sokatch

Submitted by:

Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763

Contents

Tables

Figures

Executive Summary

Background

The Analysis of Middle School Math Systems (AMS) project is part of a larger set of investments by the Bill & Melinda Gates Foundation intended to help students who are Black, Latino, multilingual learners, and/or experiencing poverty succeed in math. A core hypothesis guiding the AMS project is that teachers' use of high-quality math curricula matters for students' classroom experiences and achievement. The study is organized around five broad inquiry areas (see box). This report focuses on inquiry area 1: curricular efficacy. It presents findings on how the adoption of Eureka Math or Into Math by middle schools in one school district affected student math achievement.

Both focal curricula meet expectations for high-quality curricula, as defined by EdReports, and are therefore rated "green." Following guidance from district leadership, between the 2019–2020 and 2021–2022 school years, most middle schools in the school district switched to using either Eureka Math or Into Math. Of the 56 schools in the efficacy study, 36 schools (64 percent) adopted Eureka Math, and 16 schools (29 percent) adopted Into Math. Most of these schools switched from SpringBoard Mathematics, a "non-green" curriculum that does not meet expectations. Finally, four schools (7 percent) did not switch curricula by 2021–2022. These schools continued using either Ready Mathematics or enVision Math, which are both rated green.

AMS study inquiry areas

1. Curricular efficacy
2. Curriculum characteristics that influence instructional enactment
3. Characteristics of professional learning that supports teacher needs and effective instructional enactment
4. Adaptations in instructional enactment
5. What influences planned and unplanned adaptations in instructional enactment

We assess the effects of these curriculum switches on (1) students' math performance on the Northwest Evaluation Association Measure of Academic Progress (NWEA MAP), (2) students' math performance on the state test, and (3) students' math course grades. We use a quasi-experimental difference-in-difference analysis to examine these effects one, two, and three years after the curriculum switch occurred. The difference-in-difference analysis measures the effect of adopting a new curriculum by comparing (1) the change in trends over time for students at schools that switched to one of the two focal curricula with (2) the change in trends over time for students at schools that did not switch curricula or had not yet switched. In addition, we explore how student and teacher beliefs regarding math differ between schools that adopted Eureka Math versus Into Math, and the relationship between these beliefs and students' math achievement, using survey data from a purposely selected sample of 10 schools.

Key Findings

Switching to Eureka Math or Into Math led to increases in students' math course grades but had no detectable effect on their achievement on standardized math tests.

We find no evidence of a statistically significant impact on either NWEA MAP or state test math standardized test scores or the percentage of students who meet proficiency thresholds, on average. In contrast, switching curricula is estimated to increase students' math GPA by 0.37 points on a 4.0 scale after one year, 0.36 points after two years, and 0.33 points after three years. Likewise, switching led to the percentage of students with a passing GPA (that is, above 2.0) increasing by 12 percentage points after one, two, and three years. Applying a Bayesian interpretation to the results, we can be very sure that both curricula had positive effects on math grades; the estimated probabilities are near 100 percent.

There were few detectable differences in the effects of switching to Eureka Math compared to Into Math.

For most outcomes, there is no detectable difference between the effect of switching to Eureka Math versus Into Math. There is one exception: The effect on NWEA MAP scores after one year of implementation was larger for schools that chose Eureka Math compared to Into Math. Bayesian interpretation also suggests that switching to Eureka Math very likely led to positive effects on standardized test scores in the first year, but this probability fell considerably after each year of implementation. For example, in the first year of implementation, Eureka Math likely led to improvements in students' NWEA MAP scores (81 percent probability of an effect greater than zero), though this likelihood fell steadily to 44 percent after two years and 11 percent after three years.

Both focal curricula were relatively less effective in raising the performance of Black and Hispanic students and female students on standardized tests compared to other students.

Exploratory subgroup analyses suggest that switching to Eureka Math was less effective in raising the state test math proficiency rates of Black and Hispanic students compared to students of other races or ethnicities. Similarly, switching to Into Math was less effective in raising the state test math scores of Black and Hispanic students compared to students of other races or ethnicities, at least in the first year of implementation. We also find similar differences by gender. For example, switching to Eureka Math significantly raised the percentage of male students scoring proficient on the NWEA MAP by 2 percentage points after one and two years. Meanwhile, this effect was 3 to 4 percentage points lower for female students in each year.

Otherwise similar students had less positive beliefs about math when taught using Eureka Math than Into Math.

Even after controlling for differences in student and school characteristics, students in Eureka Math schools on average had lower scores on achievement identity, math enjoyment, math self-efficacy, and student engagement than students in Into Math schools. For example, students in Eureka Math schools scored about one point lower, on average, on the achievement identity and math self-efficacy scales, which range from 1 to 6. This difference is important because students' beliefs about their achievement identity and math self-efficacy had the strongest positive relationships to their math achievement. For instance, students with high achievement identity (or high math self-efficacy) were about twice as likely as students with low achievement identity (or low math self-efficacy) to be proficient on the state math test.

Compared to teachers in Into Math schools, teachers in Eureka Math schools were more likely to report their curriculum was coherent but less likely to feel it was appropriate for their students.

On average, teachers in Eureka Math schools had stronger perceptions of the coherence of the curriculum with content standards, assessments, and instructional policy (effect size 0.36). Teachers in both groups had similar beliefs regarding the specificity of the curricula—that is, the level and clarity of detail provided to implement it (effect size = -0.01). However, teachers in Eureka Math schools were less likely to state the curriculum was appropriate for their students (effect size = -0.19). Teachers' perceptions of their curricula were not consistently related to their students' math achievement.

Compared to teachers in Into Math schools, teachers in Eureka Math schools were more likely to report using ambitious and culturally responsive instruction and less likely to report using procedural instruction.

Teachers in Eureka Math schools were more likely than teachers in Into Math schools to report using ambitious instructional strategies requiring students to explore multiple representations of mathematics, evaluate representations and approaches to mathematical concepts or procedures, and engage critical thinking skills (effect size = 0.42), as well as using culturally responsive instructional strategies (effect size = 0.38). In addition, they were less likely to report using procedural instructional strategies that are not conceptually sophisticated (effect size = -1.01). Teachers' reported use of ambitious instruction was strongly correlated with higher student grades, though not test scores. Neither culturally responsive nor procedural instruction were correlated with student achievement.

Conclusion

Overall, switching curricula was primarily effective in raising students' math course grades, with improvements in students' math GPA occurring in every year of implementation. However, it is unclear whether these improvements reflect increases in students' math knowledge and skills or reflect changes or differences in schools' grading practices. Both Eureka Math and Into Math had similar effects on student achievement outcomes, though there was some indication that Eureka Math may have been more effective in the first year of implementation. However, results from a Bayesian analysis indicate the probability of positive effects for both curricula decreased with time after the curricula were adopted.

These results should be viewed as being specific to the place and time in which the curricula were implemented. Over half of the schools that switched curricula in the school district did so in the 2020–2021 school year—the middle of the COVID-19 pandemic. Nevertheless, we do not find evidence that the effects differ depending on the year that schools switched curricula nor the number of implementation years that had elapsed. Although all schools were impacted by the pandemic (and therefore the difference-in-difference design accounts for its overall impact), adopting a new curriculum in the period of 2020–2021 to 2021–2022 presented unique challenges that may have affected its effectiveness, even two to three years later.

Our findings, taken together with other studies of curricular efficacy that have found the same curriculum can vary in its effectiveness across contexts, suggest the curriculum that is most effective in one context may not be right in another. Research that focuses on the features of effective curricula, as well the features of teachers' enactment and the implementation context that make curricula more effective, may hold promise. This type of evidence may be more generalizable than individual efficacy studies and may help schools select curricula that are most likely to be effective in their contexts.

I. Introduction

The Analysis of Middle School Math Systems (AMS) project is part of a larger set of investments by The Bill & Melinda Gates Foundation intended to help students who are Black, Latino, multilingual learners, and/or experiencing poverty succeed in math. A core hypothesis guiding the AMS project is that teachers' use of high-quality math curricula matters (Cobb & Jackson, 2011; Elmore et al., 2014): If high-quality curricula are grounded in professional learning that aligns with the intended curricula and supports teachers to enact the curricula in a culturally responsive way, then teachers should draw from rigorous, standards-aligned, culturally responsive curricula so that students who are Black, Latino, multilingual learners, or experiencing poverty will have a better classroom experience. We are particularly interested in exploring how curricula affect the math classroom experiences of middle school students in these priority groups, which sometimes disengage from school—and from math in particular—during the middle school years in ways that have long-term implications for their academic and economic success (Balfanz & Byrnes, 2006). Therefore, a core goal of the study is to investigate whether certain curricula can make middle school math experiences more enjoyable and productive for these priority students. The study is organized around five broad inquiry areas (see Box I.1).

Box I.1. Study inquiry areas

1. Curricular efficacy
2. Curriculum characteristics that influence instructional enactment
3. Characteristics of professional learning that supports teacher needs and effective instructional enactment
4. Adaptations in instructional enactment
5. What influences planned and unplanned adaptations in instructional enactment

In this report, we focus on inquiry area 1: curricular efficacy. We present findings on how the adoption of Eureka Math or Into Math by middle schools in one school district affected student math achievement. These two curricula meet expectations for high-quality curricula, as defined by EdReports, and are therefore rated "green." Following guidance from district leadership, between the 2019–2020 and 2021–2022 school years, most middle schools in the school district switched from using a "non-green" curriculum to using either Eureka Math or Into Math. We assess how changing these curricula affected students' math standardized test scores and course grades using a quasi-experimental difference-in-difference analysis. Our assessments are made one, two, and three years after the switch occurs. The results can shed light on whether moving from a non-green to a green curriculum leads to improved math achievement, and whether one of the two green curricula adopted is more effective than the other. In addition, we explore how student and teacher beliefs regarding math differ. District leaders expressed interest in using the findings from this report to guide future curriculum policy. However, it is important to remember that this analysis assessed the effects of adopting a new curriculum during the exceptional conditions of the COVID-19 pandemic; therefore, the findings may not apply under a different, more normalized implementation context.

In the rest of this chapter, we summarize findings from the research literature on math curricular efficacy and from past AMS study reports focused on the other inquiry areas. Chapter II describes the study's context, including a description of the curricula schools switched from and to, and information on how schools chose between the two green curricula. In Chapter III, we summarize the study's data sources, analysis methods, and limitations. Chapter IV presents the findings. The report's final chapter offers a discussion of the findings given what we know about the implementation context in the school district and implications for future policy and research. Appendix A presents additional technical details about the study design, and Appendix B presents additional results. Appendix C provides analysis from an additional year of data from the school district.

Past research on math curricular efficacy

The research literature is mixed on the extent to which math curricula impact student achievement and which curricula are most effective. Research suggests that high-quality, standards-based curricula and professional learning can improve a teacher's ability to deliver more ambitious and inclusive instructional practices that incorporate diverse learners' multiple mathematical knowledge bases (students' mathematical thinking, as well as their cultural, linguistic, and community-based knowledge) into their instruction (Desimone & Garet, 2015; Moyer et al., 2011; Tarr et al., 2008; Turner et al., 2012). However, rigorous experimental and quasi-experimental studies assessing the effectiveness of specific curricula sometimes yield inconsistent results. This lack of consensus may be explained in part by the inherent difficulty in assessing the effectiveness of curricula. Mixed findings across studies could be due to differences in (1) the study designs used, (2) the math assessments used to measure student performance and how aligned they are to the curricula under study, (3) the curricula being compared and the extent to which they differ from each other, and (4) the context in which the curricula are implemented, including how they are adapted by teachers and the quality of teachers' enactment. Indeed, the same curricula are sometimes found to have varying effects on student achievement. For example, different studies of the University of Chicago School Mathematics Project (UCSMP) Algebra I curriculum have yielded a range of negative, null, and positive impacts (Slavin et al., 2008). Thus, the National Research Council (2004) recommends that a curriculum be deemed "effective" only when there are multiple rigorous studies that consistently demonstrate improvements in student learning.

With these complexities in mind, several studies demonstrate that curricula *can* make a difference for student learning. For instance, in a rigorous randomized control trial across 12 school districts, students in elementary classrooms using Math Expressions or Saxon Math outperformed those using enVision Math by 0.12 and 0.17 standard deviations, respectively (Agodini et al., 2009. Most randomized control trials are limited to a small number of districts that agree to participate in the study and may not generalize more broadly. Though relatively less rigorous, quasi-experimental studies can more readily assess curricular efficacy in larger populations. In three quasi-experimental studies that included all schools in three states (California, Florida, and Indiana), researchers identified moderate effects of some curricula on student achievement measured using state assessments (Bhatt & Koedel, 2012; Bhatt et al., 2013; Koedel et al., 2017). For example, Silver-Burdett Ginn (SBG) Mathematics and Scott Foresman‑Addison Wesley (SFAW) both outperformed Saxon Math by 0.13 and 0.06 standard deviations, respectively.

Despite evidence that specific curricula can make a difference for students in some contexts, the extent to which curricula matter for student learning, on average, is unclear. In the largest study of curricular efficacy conducted to date, which analyzed data from all schools in six states over multiple years, there was little evidence of differences in curricular efficacy (Blazar et al., 2020). Among 38 curricula, no single curriculum stood out as a consistent high or low performer across all states and years based on student achievement on state assessments. However, findings differed across states, which illustrates the importance of the study context. For example, in California, Eureka Math had a negative impact on student achievement compared to enVision Math, whereas the two curricula were found to be similar across all the states in the study. The authors conclude that adopting a new curriculum is unlikely to be an effective approach to improving student learning.

However, recent reviews of the existing literature have called for improvements in how curricular efficacy is assessed. One review by Gold et al. (2023) of 61 studies of math curricular efficacy identified several limitations to existing research. First, most studies focus narrowly on student test score outcomes. As a result, few examine other important outcomes, such as student motivation. Second, few studies examine teacher practice and perspectives or how the curricula were implemented. Third, few studies examine findings by student subgroups, such as differences by student race/ethnicity and gender. A review by Steiner (2017) identified additional limitations. For example, we know little about what makes a given curriculum effective, in terms of both the characteristics of the curriculum and the conditions under which it is implemented—including the extent to which it is implemented with fidelity. In addition, most studies take place over the course of one year or less, though it may take years for teachers to master the shift from one curriculum to another. Finally, new curricula are often compared to "business as usual," with researchers failing to clearly define what this control condition entails. Based on a review of nearly 700 studies and expert input, the National Research Council (2004) developed a Framework for Evaluating Curricular Effectiveness that addresses many of these limitations. In particular, the framework recommends using multiple methodologies to assess effectiveness by complementing rigorous, well-designed experimental or quasi-experimental comparative studies with curriculum content analyses and case studies of implementation of the curricula.

Key findings from past AMS reports

Findings from other reports in the broader AMS project provide important insights that are critical to consider when assessing the relative efficacy of curricula. In particular, other analyses have studied (1) the characteristics of curricula rated green by EdReports as compared to non-green curricula and the professional learning teachers receive, (2) teachers' instructional practices and adaptation of the curricula, (3) student beliefs regarding math and how these relate to teachers' beliefs, and (4) district and school leaders' vision for math instruction. Below, we summarize key findings from the first interim report (Amos et al., 2022a) and the second interim report (Amos et al., 2022b). These findings are based on all study curricula (see Box I.2) and districts. In addition to partnering with the school district, the AMS study team partnered with three other school districts.

Box I.2. Middle school math curricula in AMS study

Green-rated curricula

1. Illustrative Math
2. Into Math
3. Eureka Math

Non-green or non-rated curricula

4. CA Math (Glencoe)
5. Big Ideas
6. Key Elements of Mathematics Success (KEMS)

Curriculum characteristics

- On average, green curricula are more aligned with the Common Core State Standards (CCSS) than non-green curricula. Among the green-rated curricula, Eureka Math is the most aligned with the CCSS.
- Although Illustrative Math and Eureka Math are more cognitively demanding than the non-green curricula, all of the study curricula are less cognitively demanding than the CCSS recommends.
- Teachers using green curricula are significantly more likely to believe that their curricula are too cognitively demanding for their students and that they have insufficient time to reasonably cover curriculum content during the school year.
- None of the study curricula meet our Culturally Responsive Math Teaching (CRMT) tool's standard for cultural responsiveness.

Professional learning characteristics

- Teachers were most likely to receive professional learning that focuses on culturally responsive practices and analyzing student work or assessment data.
- Teachers were most likely to perceive professional learning activities as valuable to their math instruction; their understanding of how students learn math; their responsiveness to students' backgrounds, cultures, and points of view; their mindset and biases about students and setting higher expectations for all students; and strategies that improve their math instruction.
- On average across all areas of professional learning, teachers felt that the support they received only influenced their math instruction to a limited extent (including strengthening their approaches to demonstrating respect for students' cultural backgrounds), which suggests ample room for improvement.
- Less than half of teachers felt that their professional learning was aligned with feedback from observations of their teaching or connected to their daily lessons.

Teacher instructional practices and adaptations

- Across the lessons observed, teachers on average earned moderate ratings for their use of ambitious instructional practices. Teachers commonly designed lessons aligned with standards and earned high ratings for presenting mathematical concepts accurately. They were least likely to create opportunities for students to engage in problem solving or use manipulatives.
- Teachers infrequently employed strategies most commonly associated with culturally responsive teaching, including engaging students' cultural and community funds of knowledge and making interdisciplinary connections. We observed no instances of teachers empowering students to use math as a tool for social justice. However, teachers created ample opportunities for students to engage in mathematical discourse.
- Teachers predominantly delivered whole-class instruction, which suggests limited efforts to create equitable learning environments that differentiate and individualize instruction according to students' needs. In addition, teachers commonly tasked students with individual work.
- The majority of teachers reported that they made productive adaptations to their curricula, including modifying lessons to ensure a more equitable experience for their students. In particular, they said that

they differentiated instruction not only for multilingual learners and students performing below grade level but also for students performing above grade level. Only a small percentage of teachers who modified their curriculum to promote culturally responsive math teaching also reported the use of strategies that specifically leverage students' cultural and community funds of knowledge.

Student beliefs

- Students were most likely to feel that their parents think they are good at math and less likely to feel that their classmates believe so. Although a majority of students reported that their friends think that they are good at math, the following were also indicated:
  - Black and Hispanic[1] students were less likely to feel their classmates view them as good at math.
  - Female students were less likely than male students to say that their parents and classmates think they are good at math
- A majority of students indicated that they care about learning math and want to be in math class. However, the following were also indicated:
  - Black and Asian students were more likely than White or Hispanic students to say that they enjoy learning new things in math.
  - Black and Hispanic students were more likely to say that they feel frustrated in math class or often feel down.
  - Female students were less likely to report enjoying, feeling good in, or looking forward to math class and were more likely to report that they feel frustrated, worry about learning new things in math, or feel down in math class.
- Two-thirds of students indicated that they plan to continue taking math classes to prepare for college (46 percent) or because they like studying math (22 percent). Students' plans to purse math coursework in the future did not differ by race or ethnicity. However, female students were more likely to say they will keep taking math classes to prepare for college, whereas male students were more likely to say they will keep taking math classes because they like studying math.

Teacher beliefs

- Teachers' instructional self-efficacy (or belief that they have the capacity to teach effectively), confidence in teaching in culturally responsive ways, and perception that their district and school leaders are supportive did not appear to be related to students' achievement identity.
- Teachers of students less likely to indicate that they believe they can achieve in math if they work hard enough (growth mindset) were more likely to feel that their curriculum is too rigorous for their students. These teachers also tended to work in schools with more Hispanic students and students who are experiencing poverty (eligible for free and reduced-price lunch).
- Teachers who reported that their curriculum is appropriate for their students' needs (that is, not too rigorous) taught in schools with more students at or above grade-level proficiency in math.

---

[1] Throughout the report, we use the race and ethnicity labels used in the corresponding data sources.

District and school instructional vision

- Our initial look at how district and school leaders describe their respective visions for math education—and the alignment between the two—suggests that instructional leaders do not consistently share a clear and consistent vision that offers concrete guidance on instructional strategies teachers could employ to realize stated learning goals.

II. Study Context

In this chapter, we provide background information about the study's context. We begin by discussing characteristics of the school district, its vision for math learning and teaching, and the guidance and support it offered schools for math curricula and professional development. Next, we provide an overview of each curriculum included in this study. Finally, we discuss schools' curriculum selection and describe the baseline characteristics of schools that opted to adopt different curricula.

The school district

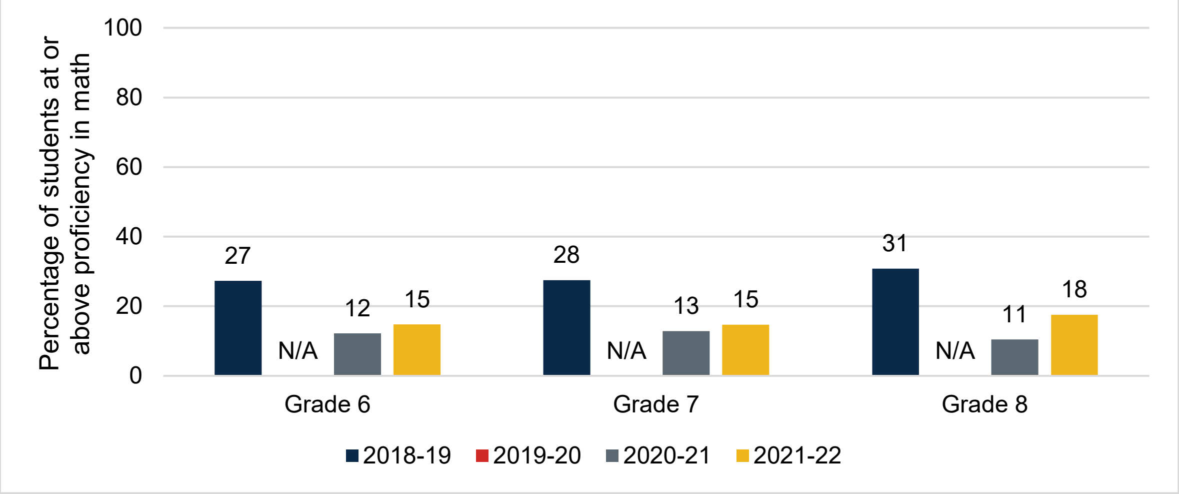
District background

With more than 35,000 K–12 students, the district is the second-largest school district in the state. The median household income is approximately $34,000 (NCES 2023). Per-student expenditures in the district are approximately $11,700, and the student–teacher ratio is just over 14 to 1.

In the 2022 state report card, the school district ranked as the state's top-rated urban district for academic growth and for closing disparities between Black, Hispanic, and other groups of students. Almost two-thirds (64 percent) of students in the district are Black, 17 percent are Hispanic, and 15 percent are White. English language learners make up 13 percent of the school district's students, and 23 percent of the district's students have special education needs. All students in the district are considered economically disadvantaged.

In the 2018–2019 school year, 29 percent of students in grades 6 to 8 scored at or above the state's proficiency level in math, as measured by the state test (Figure II.1). In 2020–2021, following the COVID-19 pandemic, the percentage of students scoring at or above proficiency in math dropped to 12 percent. Although proficiency increased in the 2021–2022 school year, it remains below pre-pandemic levels: Just 16 percent of the district's middle school students scored at or above proficiency last year.

Figure II.1. Students in grades 6 to 8 scoring at or above the state's proficiency level in math



Source: State Department of Education data.

Note: The state test was cancelled in the 2019–2020 school year due to the COVID-19 pandemic.

N/A = not available.

District math vision

As described in its flyer on math learning and teaching priorities and strategies for the 2019–2020 school year, the school district envisions that students will engage in cognitively challenging tasks and that every student will be a confident and flexible mathematical thinker, communicator and problem-solver. To enact this vision, the district plans to ensure students have access to high-quality math instruction, regular monitoring of progress, and extra help as needed. The district aims to provide differentiated support to students when implementing grade-level tasks and aims to provide teachers with research-based strategies for math instruction. Teachers are encouraged to follow guidance from the Math Scope & Sequence document, developed by the National Council of Teachers of Mathematics (NCTM), which contains research-informed practices that promote effective math teaching and student learning. Research-based practices include establishing math goals, using various representations of concepts, facilitating meaningful math conversations in which students can share ideas, and allowing students time to struggle with ideas and make sense of new concepts.

To help schools enact this vision, each year schools must write an academic achievement plan that outlines their instructional strategy, including for math. Math strategies must be aligned with the NCTM's Mathematics Teaching Practices. District staff meet with principals about their plan each month to provide implementation support. In monthly 90-minute coaching sessions, principals receive coaching on implementing one of the following NCTM principles: Use and connect mathematical representations, facilitate meaningful mathematical discourse, or implement tasks that promote reasoning and problem-solving. District instructional leadership noted that "for many principals, math can be intimidating, so these sessions are focused on helping principals have tools to support their teachers."

District math curricula and professional learning supports

Starting in the 2019–2020 school year, the school district's leadership provided updated guidance to schools for their math curriculum selection. Schools were given two options vetted by the district: Eureka Math and Into Math. Although schools were not required to adopt either one, schools that selected one of the district-vetted options did not need to pay for the curriculum or for professional development, as these would be covered by the district. Interviews with district instructional leadership suggest Eureka Math was the first curriculum they selected. Many schools and teachers in the school district had already expressed interest in the curriculum, which began as EngageNY, a free and open education resource with "high-quality resources [and] high level of rigor and expectations." Into Math was then selected that year to provide schools with an alternative to Eureka Math, which some schools had already been piloting before 2020: "With Eureka, we found that teachers weren't doing enough advance planning ... Into Math is a more traditional type of curriculum (more scripted for teachers), so we wanted to offer that option."[2] The district invited curriculum developers for Eureka Math and Into Math to give presentations to school leaders about the curricula to inform their decision.

Two professional learning providers, Houghton Mifflin Harcourt and Teaching Lab, offered professional

---

[2] Starting in the 2023–2024 school year, the district will also give schools the option to move from Eureka Math to Eureka Math Squared, which offers an enhanced online platform and more streamlined materials.

development supports to teachers.[3] In addition, the district's Model Math Lead Teacher offers professional learning support to teachers and principals, including regular office hours and ad hoc sessions by request from schools. For each curriculum, the district offers a menu of professional learning options. However, the district does not mandate how much professional learning teachers must receive; it is up to school leaders to make these plans. In practice, district instructional leadership noted that "schools have not always done their due diligence to schedule it … there is a lot of variation in how (and how often) principals scheduled [professional development] sessions." District instructional leadership also noted that the COVID-19 pandemic made professional learning more challenging. Teachers burnt out from virtual sessions during the peak of the pandemic, but because the district has a shortage of substitute teachers, it has been difficult to offer in-person sessions. Providing sessions before or after school has also been challenging, as there are four different school schedules in the district.

Overview of the study curricula

Of the 33 study schools that switched curricula in the 2020–2021 or 2021–2022 school years, the vast majority (28 of 33, or 85 percent), switched from SpringBoard Mathematics.[4] Additionally, four schools switched from enVision Math and one switched from Ready Mathematics. Of the 56 schools in the efficacy study, 36 schools (64 percent) adopted Eureka Math and 16 schools (29 percent) adopted Into Math. Finally, four schools (7 percent) did not switch curricula by 2021–2022. Table II.1 summarizes the curricula schools used before and after switching. Below, we provide an overview of each curriculum.

Table II.1. Curricula study schools switched from and to, 2019–2020 to 2021–2022

Previous curriculum	Number of schools that switched to Eureka Math	Number of schools that switched to Into Math	Number of schools that did not switch by 2021–2022	Total
enVision Math	2	2	1	5
Ready Mathematics	1		3	4
SpringBoard Mathematics	14	14		28
Unknown	19			19
Total	36	16	4	56

Source: School district data.

Note: Data on previous curriculum use was not available for the 19 schools that adopted Eureka Math in 2019–2020.

---

[3] Teaching Lab provided supports to teachers participating in the AMS study.

[4] Data on schools' curriculum use were only available for the 2019–2020, 2020–2021, and 2021–2022 school years. Therefore, we are unable to determine which curricula the 19 schools that switched in 2019–2020 had used prior to 2019–2020.

Into Math

Into Math is a K–8 core curriculum published by Houghton Mifflin Harcourt. The curriculum is print based and offers digital and interactive versions of many curriculum components. Online resources include data dashboards that help teachers monitor student learning, identify student learning needs, and guide teachers on how to differentiate instruction to meet student needs. Into Math emphasizes conceptual understanding and reasoning skills first, before connecting students' understanding to procedural practice on concepts and skills. Each lesson begins with a whole-class activity to introduce the day's task and build understanding through student discussion. Teachers then lead a five- to 10-minute whole-group lesson that connects mathematical concepts and skills to the understanding developed in the initial activity. Students then engage in a quick formative assessment, which enables teachers to check understanding. Students then receive differentiated instruction through small groups, independent practice, or enrichment activities. All lessons end with an activity where the teacher brings students together to summarize what they learned.

Reviews of the curriculum note the following:[5]

- It supports the vertical alignment of curriculum standards from grade to grade, supports student conceptual understanding, and supports students in applying math concepts to real-world experiences.
- It encourages multiple representations when solving problems and student discourse; students are asked to justify their solutions and explain their thinking to others and evaluate others' thinking.
- The materials support teachers in meeting diverse student needs through scaffolds, differentiating tasks, and language resources, though they may need to be adapted to draw upon students' diverse cultural backgrounds.
- The materials are designed to help teachers create grouping strategies based on student readiness to support differentiated instruction, and they include informal and formal assessments to measure and adjust instruction; however, they lack resources for building teachers' content knowledge.

Eureka Math

Eureka Math—also known as EngageNY—is a curriculum for grades pre-K to 12 that sequences mathematical progressions to present high-quality math content, teach persistence in problem solving, and prepare students to understand advanced math. The curriculum emphasizes mathematical thinking and problem solving and understanding of math concepts so students can use numerous strategies for solving difficult problems. Eureka Math was developed to align with and meet the CCSS standards. It provides comprehensive print and digital curriculum materials and professional development. Lessons support classroom routines; each lesson includes discussion, exercises, a closing, and exit tickets. The entire pre-K to 12 Eureka Math curriculum, along with a variety of instructional materials and support resources, can be downloaded at no charge. Some materials, such as printed workbooks, the Eureka Digital Suite, and Affirm (the formative assessment package) require a fee.

---

[5] See https://www.doe.mass.edu/rlo/instruction/curate/index.html#/lessons/tB0yzumXkCNGduk2nfBwFDjn7Z6BAat9.

Reviews of the curriculum note the following:

- The materials provide an effective progression from conceptual to abstract, with a strong mathematical foundation; the materials provide explanations to help build teachers' content knowledge.
- The materials provide frequent opportunities for real-world application, and they ask students to justify their reasoning and critique the reasoning of others; however, they are often text heavy and can be difficult for struggling readers or multilingual learners.
- The materials prompt more teacher-led dialogue than peer-to-peer conversation, require teachers to supplement or adapt significantly to meet the needs of all students, and move quickly through the content; additionally, there are no strategies or assessments for assessing students' prior knowledge.

enVision Math

enVision Math is a K–12 curriculum published by Savvas Learning Company (formerly Pearson). The curriculum is accompanied by Savvas Realize®, which is a learning management system that includes all enVision digital math content for students, assessments, student data, pre-built lesson presentation slides, and management tools. Students can access content offline. The platform enables teachers to customize lessons and integrate with other platforms (such as Canvas®, Schoology®, and Google Classroom®). The materials also include family engagement resources that provide information for families to use in supporting their students at home (and can be accessed regardless of language spoken at home, using Google Translate; materials are available in Spanish).

The curriculum is designed to help students develop an understanding of math concepts through problem-based instruction, small-group collaboration, and visual learning with a focus on mathematical reasoning and modeling. Lessons are structured around an approach called 3-Act Math Modeling, designed to help students think mathematically. Act 1 is The Hook, which involves a video to engage students and prompt brainstorming in small groups about predictions; Act 2 is The Model, in which students identify important information and develop mathematical models to arrive at a solution; and Act 3 is The Resolution, in which students explain differences between their own conjectures and the actual solution. Differentiated instruction and ongoing assessment are used to meet the needs of students at all ability levels. The materials also focus on comprehensive vertical alignment across the grade levels.

Reviews of the curriculum note the following:

- Materials balance conceptual understanding, procedural fluency, and real-world application, though teachers need to supplement the materials to ensure students' diverse identities are affirmed in instruction. (Materials affirm diverse backgrounds at a superficial level, such as using names traditionally associated with non-White communities.)
- Materials encourage students to use multiple representations and collaborate with peers but do not explicitly encourage students to explain their thinking or evaluate others'; also, they encourage the use of a range of tools but lack guidance on how to use them.
- Teachers are provided good guidance on how to support the needs of multilingual learners; there are also teacher-focused content explanations for each topic.
- The materials include many opportunities for formal and informal assessments.

- Small group instruction is often suggested, but there is no guidance on what teachers should consider in determining when to use this instructional approach.

Ready Mathematics

Ready Mathematics is a curriculum for grades K–8 published by Curriculum Associates. It includes printed and online resources and professional learning resources that are built into the beginning of every unit and embedded within every lesson. It also offers professional guidance to support multilingual students, including language differentiation strategies that provide scaffolds and ideas for community and cultural responsiveness that can be incorporated during the lesson. Ready Mathematics emphasizes conceptual understanding through reasoning, modeling, and discussion. It places a strong emphasis on mathematical discourse and the integration of a Think-Share-Compare instructional routine, which aims to have students lead the majority of classroom discussion. It is designed to build students' confidence to problem solve, develop productive habits, and enable teachers to engage students in rich mathematical discussions. There are also frequent opportunities for practice and assessment to ensure that students understand concepts and to help teachers make instructional decisions.

Reviews of the curriculum note the following:

- Although the materials provide good connections across grade levels, they miss opportunities for connecting content across domains; teachers may need to provide supplemental materials to address this.
- Students are encouraged to use a range of mathematical tools, engage in peer-to-peer discourse, investigate classmates’ thinking, and demonstrate their knowledge.
- There are recommendations for supporting multilingual learners but not always clear guidance on how to engage students with different levels of language acquisition.
- The lessons rely on a single routine that isn’t always clear, and there are not always enough teacher supports.

SpringBoard Mathematics

SpringBoard Mathematics is a curriculum for grades 6 to 12 published by the College Board. It offers both print and digital core instructional materials. Teachers have access to professional learning that ranges from e-learning modules to multi-day workshops. SpringBoard is structured using the Understanding by Design instructional model. Students “get ready” for learning by examining the types of assessment questions they will experience when instruction is complete. (According to the publisher, the curriculum focuses on building math skills to prepare students to do well on assessments such as the PSAT 8/9). Lessons balance conceptual understanding and procedural fluency and provide opportunities for individual and collaborative learning experiences. Problems are framed in terms of meaningful, real-world applications, and examples provide step-by-step guidance toward solutions. The complexity of problems increases progressively, enabling students to deepen and extend their learning. Each unit contains discussion group tips to encourage collaboration, mini lessons that review prior concepts, formative and summative assessments, and differentiated instruction to address the needs of both struggling and advanced students. Students have additional opportunities to work in pairs or small groups by completing digital work through SpringBoard Digital, which is based on a partnership with Desmos.[6]

Reviews of the curriculum note the following:

- There are too few days allocated to the major content for each grade level; there needs to be additional material to ensure students grasp the content for each grade level.
- Content does not highlight when material from prior grades is being reviewed or when material will be relevant for future grades; the materials do not make these connections in ways that will prepare students for upcoming grades.
- Materials give students ample opportunities to work on grade-level problems, though they do not always address the needs of students with varying levels of ability; also, the materials do little to guide teachers in differentiating instruction for students.
- Teachers need to make significant modifications to the materials to cover content and make appropriate connections.

Comparing the curricula

To compare the study curricula, we used data from EdReports, which issues overall alignment ratings of "meets, partially meets, or does not meet expectations" (which are color coded as green, yellow, and red, respectively). Ratings are determined by how much instructional materials align with grade-level learning and mathematical practice standards, facilitate student learning, and enhance a teacher's ability to differentiate and build knowledge within the classroom (Box II.1). Educator-led review teams review the curricula independently and then discuss evidence as a team to issue the ratings. We compiled ratings from the publicly available EdReports Report Database.

Four of the five study curricula earned a green rating from EdReports, whereas SpringBoard Mathematics—the curriculum most used before schools switched to Eureka Math or Into Math—was not rated green. Table II.2 presents the ratings for each curriculum, by domain and grade level (6 to 8). The most notable differences are between the green curricula and the non-green curriculum (SpringBoard Mathematics), which generally does not meet expectations. Green curricula primarily diverge in their scores on usability.[7] Into Math, enVision Math, and Ready Mathematics scored above 90 percent for usability, whereas Eureka Math received a score of 66 percent.

Box II.1. EdReports rating domains

EdReports issues a meets (green), partially meets (yellow), or does not meet expectations (red) overall alignment rating based on a curriculum's performance across three domains:

1. **Focus & coherence:** Do the materials assess grade-level content, and are they coherent and consistent with the Common Core State Standards (CCSS)?
2. **Rigor & mathematical practices:** Do the materials meet the CCSS expectations for rigor and mathematical practices?
3. **Usability:** Do the materials support teachers to fully utilize the curriculum, understand the skills and learning of their students, and support a range of learners?

---

[6] See https://www.learninglist.com/reviews-college-boards-SpringBoard-mathematics/.

[7] Non-green curricula are not scored on their usability.

Table II.2. EdReports ratings of study curricula

	Green curricula				Non-green curricula
	Into Math	**Eureka Math**	**enVision Math**	**Ready Mathematics**	**SpringBoard Mathematics**
Overall alignment rating	Meets Meets Meets	Meets Meets Meets	Meets Meets Meets	Meets Meets Meets	DNM DNM Partial
Focus & coherence	14/14 14/14 14/14	14/14 14/14 14/14	14/14 14/14 14/14	13/14 14/14 14/14	6/14 6/14 11/14
Rigor & mathematical practices	17/18 17/18 17/18	16/18 16/18 16/18	17/18 17/18 17/18	17/18 16/18 17/18	N/A N/A 9/18
Usability	35/38 35/38 35/38	25/38 25/38 25/38	38/38 38/38 38/38	36/38 36/38 36/38	N/A N/A N/A

Source: EdReports data.

Note: Each grade is reported separately for each curriculum; grade 6 is the first set of numbers in each row, and grade 8 is the last set of numbers in each row. Color coding matches EdReports ratings. Ratings of enVision Math are for the 2021 version. Earlier versions, which some study schools may have used, were not rated by EdReports.

DNM = does not meet; N/A = not applicable.

For the two curricula that are the focus of this study—Eureka Math and Into Math—the AMS project gathered additional data from two sources: the Surveys of Enacted Curricula (SEC) and the Culturally Responsive Mathematics Teaching (CRMT) curriculum coding tool. The SEC assesses the extent to which a curriculum's topic emphasis and cognitive demand of student performance expectations aligns with the CCSS.[8] The CRMT curriculum coding tool measures the prevalence of guidance in a curriculum to implement culturally responsive instruction, such as connecting content to student culture and identities, providing all students with rigorous material,[9] and attending to the power and participation of students throughout the learning process. The AMS team developed the tool by using an adaptation of the CRMT Lesson Analysis Tool (Aguirre & del Rosario Zavala, forthcoming).[10]

---

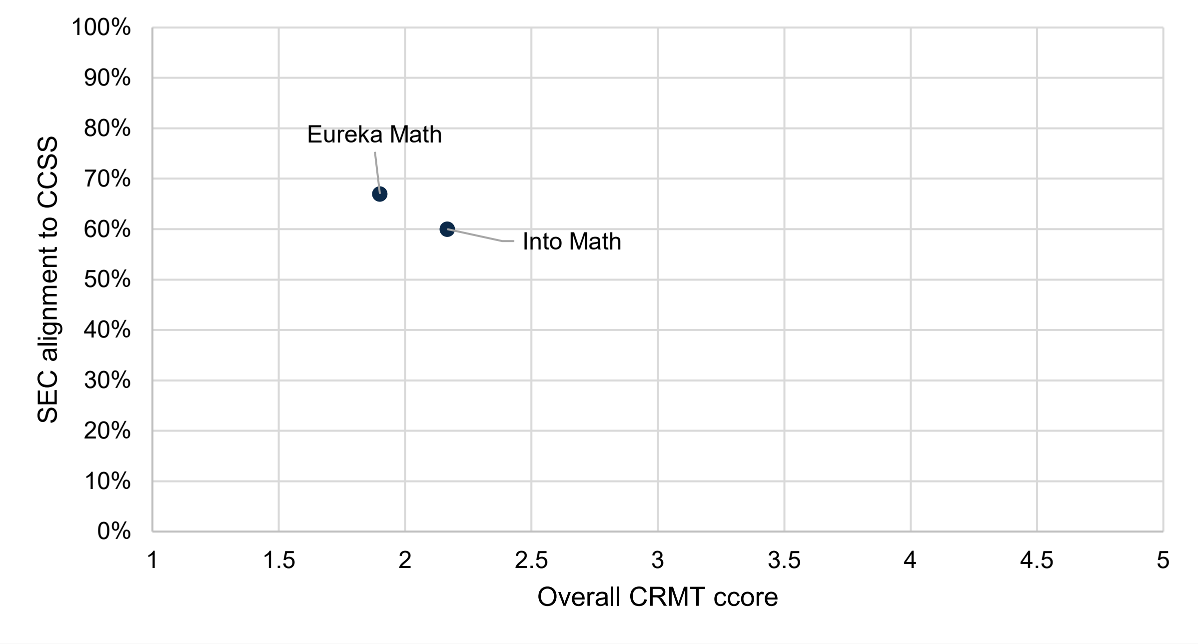
[8] A team of math education experts with the Center for Curriculum Analysis (CCA) scored the study curricula using this tool. Scores represent the full middle school curriculum (grades 6 to 8).

[9] Although including rigorous material is not unique to CRMT, our CRMT coding tool uniquely measures when guidance exists to ensure *all* students have access to rigorous material. EdReports and the SEC measure whether more rigorous material exists, but not which students the curricula suggest access each task.

[10] The tool was adapted because the original tool was designed to be used in dialogue with teachers rather than as a means of quantitatively scoring materials. A team of trained Mathematica staff reviewed the teacher's guide for each curriculum independently and noted evidence or opportunities to enact CRMT within each domain. After coding, the review team came to a consensus on discrepancies in the codes before finalizing the data. Unlike EdReports and the SEC—which issues ratings based on the full curriculum for grades 6 to 8—the CRMT coding tool review team reviewed a sample of 6th grade lessons (i.e., six to nine lessons across four units) to issue CRMT scores.

Although Eureka Math is more aligned to CCSS than Into Math (according to the SEC), Into Math received a higher overall rating on cultural responsiveness on the CRMT. Both curricula received a high alignment score (60 percent for Into Math and 67 percent for Eureka Math). However, neither curriculum received a high score on cultural responsiveness. On the CRMT's scale of 1 to 5, Into Math received a score of 2.2, and Eureka Math scored 1.9. Figure II.2 summarizes the results of these assessments. For additional detailed results from the SEC and CRMT analyses of the AMS study curricula, see the report "Analyzing Middle School Math Curricula: A Comparative Study Using Three Measurement Tools" (Stone et al., 2023).

Figure II.2. SEC and CRMT ratings of Eureka Math and Into Math



Source: Center for Curriculum Analysis and Mathematica data.

CCSS = Common Core State Standards; CRMT = Culturally Responsive Mathematics Teaching; SEC = Surveys of Enacted Curricula.

Schools' curriculum selection

Because this is not an experimental study in which curricula are randomly assigned to schools, it is important to understand how schools decided which curriculum to adopt and, therefore, how schools selecting different curricula might differ. Interviews of 11 principals in the school district offer some insight into the decision. Of the nine principals who gave reasons for their school's choice, the most common reason cited (by three principals) was hearing positive things about the curriculum from other teachers and school leaders; they gave this reason exclusively for Eureka Math, which had been piloted by some teachers and schools before the district issued the new guidance. For example, one principal noted they "heard from other schools that they liked Eureka, and so, it was kind of word of mouth." The next most common reason was preferring the materials or approach offered by one curriculum over another; schools choosing both Eureka Math and Into Math gave this reason. For example, one principal stated the following:

> My teachers are not ready to do Eureka. That would have not been a successful program here. Into Math was much more of a reality for the staff that I have…Some of my friends and our principals use Eureka and their teachers were struggling because it does change your mindset on how you introduce math and how you do math. And I know my staff … they do not come easy to change. … So, Into Math was the better next step for this group of people. Don't get me wrong, Into Math is challenging, and I actually think it's more challenging than they all thought when they just previewed the program.
>
> - School principal

When comparing the characteristics of schools that made different curricular choices, we find several differences between Eureka Math and Into Math schools (Table II.3). Schools that adopted Eureka Math had lower percentages of Black students than schools that adopted Into Math (66 versus 81 percent) and higher percentages of Hispanic students (15 versus 7 percent), White students (15 versus 9 percent), and English language learners (5.3 versus 2.5 percent). Before adopting the new curriculum, schools that switched to Eureka Math had somewhat higher shares of students who scored at or above proficient in math on the state test (20 versus 16 percent) and received passing grades in math courses (50 versus 42 percent). Whereas schools that adopted Into Math did not enroll any students in math intervention classes, about 7 percent of students in Eureka Math schools were enrolled in math intervention classes.[11]

Schools that did not switch curricula by 2021–2022 also differed from schools that switched along some dimensions: Compared to schools that switched, they had higher shares of Hispanic students (40 percent) and English language learners (15 percent). Students in schools that did not switch curricula tended to have similar levels of math proficiency as students in Eureka Math schools. On the state test, 20 percent of students in schools that did not switch scored at or above proficient in math. These schools also enrolled a much higher percentage of students in math intervention classes (42 percent).

---

[11] Over 99 percent of students in each curriculum school group were enrolled in grade-level math courses, which are the focus of this study.

Table II.3. Baseline characteristics of schools, by curriculum choice

	Eureka Math schools (N = 36)		Into Math schools (N = 16)		Other schools (N = 4)		Standardized differences		
	Percent or mean	SD	Percent or mean	SD	Percent or mean	SD	Eureka Math vs. Into Math	Eureka Math vs. other	Into Math vs. other
English learner (%)	5.3	11.5	2.5	4.4	14.6	18.1	0.48	-0.67	-1.15*
Special education (%)	25.9	7.5	25.2	7.1	24.2	2.6	0.02	0.06	0.03
Black (%)	65.8	30.2	80.5	27.9	55.4	49.9	-0.46	0.26	0.73
Hispanic (%)	14.8	16.9	7.4	11.7	40.3	45.5	0.47	-0.82*	-1.29*
White (%)	15.1	15.6	9.4	15.7	3.7	4.5	0.33	0.94	0.61
Female (%)	46.8	15.7	50.0	14.2	46.9	2.3	-0.08	0.00	0.08
State test math z-score (mean)	0.02	0.42	-0.11	0.40	-0.13	0.59	0.30	0.33	0.04
Proficient state test math (%)	20.4	15.9	16.4	14.3	20.3	17.9	0.16	0.00	-0.16
NWEA MAP math z-score (mean)	0.05	0.34	-0.03	0.35	-0.19	0.37	0.24	0.69	0.42
Proficient NWEA MAP math (%)	5.0	5.4	4.3	5.3	5.0	4.8	0.08	-0.01	-0.09
Math course GPA (mean)	2.16	0.39	1.93	0.44	1.92	0.77	0.55	0.53	0.01
Math course passing grade (%)	50.4	14.1	42.0	14.4	43.9	23.2	0.20	0.16	-0.05
Enrolled in math intervention course (%)	7.1	20.0	0.0	0.1	41.5	48.7	3.23	-1.35**	-4.58**
Enrolled in advanced math course (%)	1.2	2.8	0.2	0.7	0.5	0.6	1.14	0.53	-0.61
Chronically absent (%)	34.4	12.6	33.8	14.0	31.1	17.2	0.02	0.09	0.08
Suspended (%)	26.6	13.9	23.3	13.2	23.7	10.5	0.11	0.09	-0.01

Source: School district data.

Notes: For schools that adopted Eureka Math or Into Math, baseline data typically come from the school year before they adopted the new curriculum (2018–2019, 2019–2020, or 2020–2021), with the following exception: Because the state test was canceled in 2019–2020, state test baseline data come from 2018–2019 for schools that switched in 2020–2021. For schools that did not switch curricula, all baseline data come from 2018–2019. State test and NWEA scale scores were converted to z-scores using district means and standard deviations by assessment, subject, grade, and year. Standardized differences between schools were computed using Hedges' g for continuous variables and Cox's Index for binary variables. SD = standard deviation.

* Difference is statistically significant at the 5 percent level.

** Difference is statistically significant at the 1 percent level.

There are also differences in the characteristics of schools that were early adopters, with students in these schools demonstrating higher rates of math proficiency than students in schools that switched curricula later, after the district issued its guidance. As noted earlier, in the 2019–2020 school year, 19 schools became early adopters of Eureka Math. Twenty-seven percent of students in these schools were proficient in math on the state test before the curriculum switch occurred, compared to about 14 percent in schools that switched to Eureka Math in later years. Similarly, schools that adopted Into Math in 2020–2021 had higher math proficiency rates than those that switched in 2021–2022 (21 versus 3 percent). These comparisons suggest that both the choice of curriculum and the timing of the decision to switch is correlated with schools' historical math performance. In the next chapter, we discuss our approach for estimating the effects of adopting Eureka Math or Into Math despite these differences.

III. Study Design

In this chapter, we provide an overview of the study's design. We begin by describing the data sources and samples used in the analyses. Next, we summarize the analytic approaches used to assess the effects of adopting Eureka Math or Into Math and to conduct additional exploratory analyses. The chapter closes with a discussion of the study's limitations. Additional details on the study's design appear in Appendix A.

Data sources and samples

This study drew on a combination of administrative, survey, and interview data available for different samples of schools in the school district. For the main efficacy analysis assessing the effects of Eureka Math and Into Math, we relied on administrative records from the school district, which were available for all middle schools in the district and spanned multiple years. For additional contextual and exploratory analyses, we used survey and interview data that were collected from a purposefully selected sample of schools in 2021–2022 as part of the AMS project. Below, we describe each data source and sample used in this study.

Administrative data from all middle schools

The school district provided student-level records for all students in grades 6 to 8 for the school years 2015–2016 to 2021–2022. These records included information on students' background characteristics and math achievement, including their scores on the NWEA MAP and the state test and their letter grades in math courses. In addition to the student-level data, the school district provided information on the math curricula used by each middle school during the 2019–2020, 2021–2021, and 2021–2022 school years—the years in which most schools switched to Eureka Math or Into Math following district guidance. Information on curriculum use in earlier years was not available. Table III.1 provides details on the types of administrative data used for the study.

Although the school district provided data for all 64 middle schools in the district, the efficacy analysis included 56 schools. We excluded one remote middle school; five schools that opened during the study period and therefore were missing some or all years of baseline data needed for the efficacy analysis; and two schools that used both focal curricula during the study period (one school switched from Eureka Math to Into Math after two years, and one school switched from Into Math to Eureka Math after one year).

Table III.1. Administrative data collected for the efficacy study

Data	Years available	Timing of data collected	Details on data constructs
Student NWEA MAP math scores	2015–2016 to 2021–2022	The test was administered in the winter of each academic year. There were no disruptions in test administration due to the COVID-19 pandemic.	We converted scale scores into z-scores by subtracting the average scale score for students in the school district within the same school year and grade level and dividing by the corresponding standard deviation. A student was deemed proficient if they met or exceeded the proficiency benchmarks recommended by the NWEA (Tran et al., 2022).

Data	Years available	Timing of data collected	Details on data constructs
Student state test math scores	2015–2016 to 2018–2019 and 2020–2021 to 2021–2022	The test was administered in the spring of each academic year, except in spring 2020, when the test was canceled due to the COVID-19 pandemic.	We converted scale scores into z-scores by subtracting the average scale score for students in the school district within the same school year and grade level and dividing by the corresponding standard deviation. A student was deemed proficient if they met or exceeded the proficiency benchmarks established by the state.
Student math course grades	2015–2016 to 2021–2022	Grades were reported quarterly for each academic year.	We converted letter grades to a four-point numeric scale (A = 4, B = 3, C = 2, D = 1, F = 0). We then averaged quarterly grades within each course to obtain an overall course grade. If a student had fewer than two quarterly grades, they were deemed to be missing a final grade. If a student's course grade was greater or equal to 2.0, they were deemed to have a passing grade. Grades on intervention and advanced math courses were not included in the analysis. See Table II.3 for details on the share of students in these courses.
Student background characteristics	2015–2016 to 2021–2022	Characteristics were reported yearly for each academic year.	Characteristics reported include the following: • Race or ethnicity • Gender • English language learner status • Special education status • Number of suspensions and expulsions, which we used to determine whether a student was ever suspended in a given year • Attendance rate, which we used to determine whether a student was chronically absent (missed 10 percent or more of school days) in a given year • State test English language arts scale score • NWEA MAP reading scale score
School curriculum use	2019–2020 to 2021–2022	Curriculum use was reported yearly for each academic year.	We used yearly data on the math curriculum used by each middle school to determine whether and when schools adopted a new curriculum, which curriculum they adopted, and which curriculum they switched from.

Survey and interview data from a subset of middle schools

As part of the broader AMS project, 10 middle schools in the district participated in a more extensive data collection effort during the 2021–2022 school year to understand curricula enactment: how and to what extent district and school instructional guidance and professional supports are aligned, and the relationships between professional supports, local school contexts, teacher knowledge and beliefs about mathematics instruction, classroom practice, and students' classroom experiences. The project team worked with district staff to identify and recruit these schools. Only schools that adopted a focal curriculum (Eureka Math or Into Math, in the case of this district) were recruited. Data collected in these schools and used in this analysis included interviews with district administrators and school principals and teacher and student surveys. Table III.2 describes each of these efforts. Additional details about how this sample of schools compares to all middle schools in the efficacy analysis, how teachers and students were sampled, survey response rates, and survey constructs appear in Appendix A.

We used the additional data available for this subset of schools to describe the context for the adoption of Eureka Math or Into Math and to explore relationships between the curricula, students' and teachers' beliefs, and students' math achievement. Because schools that participated in additional data collection were surveyed after adopting one of the focal curricula, it was not possible to assess changes in students' and teachers' beliefs before and after switching to Eureka Math or Into Math. In addition, the students and teachers involved in additional data collection are not representative of the district. Therefore, analyses of these data are seen as contextual or exploratory.

Table III.2. Additional data collected for the enactment study

Data	Years available	Timing of data collected	Details on data constructs
District administrator interview	2021–2022	Conducted in the fall	Interview topics included approaches to selecting and implementing math curricula, the district's vision for high-quality math instruction, and professional learning supports provided for math teachers.
School administrator interview[a]	2021–2022	Conducted in the fall	Interview topics included vision for high-quality math instruction, approach to providing instructional leadership and selecting a curriculum, and the professional learning supports, including coaching, provided to math teachers in their school.
Teacher background survey	2021–2022	Administered in the fall	This survey asked about teaching background and experience, perceptions and use of various math curricula, and teaching practices.
Teacher follow-up survey	2021–2022	Administered in the spring	This survey asked about professional support, teacher beliefs, enactment and adaptation of curriculum planning, and instructional delivery.
Student survey	2021–2022	Administered in the fall and spring	This survey asked about students' math achievement identity, growth mindset, enjoyment, self-efficacy, and engagement.

[a] In a few cases, the principal delegated math instructional leadership to an assistant principal or instructional coach and requested that we interview this person instead.

Overview of analysis methods

To assess the effects on student achievement of adopting Eureka Math or Into Math, we used a longitudinal quasi-experimental research design known as difference-in-difference. We also conducted additional analyses to help us interpret the main findings. First, we used a Bayesian approach to interpret the main findings and determine the likelihood of positive effects. Second, we assessed whether the effects of the curricula differed for subgroups of students. Third, we assessed whether the main findings were robust to different analytic decisions. Finally, we conducted descriptive analyses of the teacher and student survey data to contextualize the main findings. Below, we describe each analytic approach.

Main analytic approach

The study sought to assess the effect of schools adopting Eureka Math or Into Math on students' achievement in the NWEA MAP, state test, and math courses (see Table III.3). To do so, we used a difference-in-difference strategy, which measures the effect of adopting a new curriculum by comparing (1) the change in trends over time for students at schools that switched to one of the two focal curricula with (2) the change in trends over time for students at schools that did not switch curricula (or had not yet switched). Even though schools that switched to Eureka Math or Into Math are likely to be different, on average, along many observed and unobserved measures compared to schools that switched later or did not switch curricula, the difference-in-difference strategy can still accurately estimate the effect of switching curricula, as long as these schools showed similar average trends in their math achievement outcomes in the years before schools switched (Angrist & Pischke, 2008; Imbens & Wooldridge, 2009).[12]

Table III.3. Outcome measures used in the efficacy analysis

Outcome	Description	Primary measures	Additional measures
NWEA MAP	An interim, state-aligned adaptive assessment designed to measure student achievement and growth over time	The student's z-score on the math test, measured in standard deviation units	A binary measure of whether the student met or exceeded the proficiency benchmark in math
State test	A state-developed summative assessment designed to measure student achievement based on the state's learning standards	The student's z-score on the math test, measured in standard deviation units	A binary measure of whether the student met or exceeded the proficiency benchmark in math
Course grades	Teacher-reported grades of student achievement in grade-level math courses	The student's math course grade, measured in GPA units (0 to 4)	A binary measure of whether the student had a passing GPA (greater than or equal to 2.0)

Note: The state test was cancelled in the 2019–2020 school year due to the COVID-19 pandemic, so the analyses exclude state test math outcomes for that year. The 2019–2020 school year is both an outcome year for schools that switched curricula in 2019–2020 and a baseline year for schools that switched curricula in 2020–2021 or 2021–2022. Although we report findings on math achievement on the state test, these findings should be interpreted with caution, as the analysis cannot account for how trends might have changed in that year.

Because schools switched curricula over the course of a three-year period (2019–2020 to 2021–2022), the comparison group of schools used to estimate these effects can be a mix of schools that never switched curricula during this period and schools that had yet to switch at the time when the outcome was measured. For example, the estimated effect of switching to Eureka Math after one year is the result of averaging the following three types of comparisons:

1. The difference in trends between students at schools that switched to Eureka Math in the 2019–2020 school year compared to a mix of schools that never switched, schools that switched to Eureka Math or Into Math in 2020–2021, and schools that switched to Eureka Math or Into Math in 2021–2022—

---

[12] Given that schools adopted new curricula on a rolling, or staggered, basis (that is, switches occurred over a three-year period rather than all at once), an additional requirement for the difference-in-difference analysis to yield unbiased estimates is that the timing of adoption is not related to differential trends in the outcomes of interest (Baker et al., 2022). Although "early adopter" schools had higher baseline math achievement than schools that switched curricula later, we find that their baseline *trends* were similar. As described in this chapter, we also test for the sensitivity of the results to the timing of the curriculum switch. See Appendix B for additional results.

using outcome data from the end of the 2019–2020 school year (the first implementation year for the 2019–2020 switchers)

2. The difference in trends between students at schools that switched to Eureka Math in the 2020–2021 school year compared to a mix of students at schools that never switched and those at schools that switched to Eureka Math or Into Math in 2021–2022—using outcome data from the end of the 2020–2021 school year (the first implementation year for the 2020–2021 switchers)
3. The difference in trends between students at schools that switched to Eureka Math in the 2021–2022 school year compared to students at schools that never switched—using outcome data from the end of the 2021–2022 school year (the first implementation year for the 2021–2022 switchers)

These comparisons can best be summarized as measuring the average effect of switching to Eureka Math or Into Math versus not switching, based on schools that never switched and schools that had not yet switched. Table III.4 shows the number of schools that switched to each curriculum by year and the number of schools that did not switch. To understand the comparisons being made, it is also important to note which curricula schools switched from and which curricula schools that never switched continued to use. As noted in Table II.1, most schools that adopted Eureka Math or Into Math (85 percent) switched from SpringBoard Mathematics, a non-green curriculum.[13] The four schools that did not adopt switch curricula continued using enVision Math (N = 1) or Ready Mathematics (N = 3), which are both rated green. Therefore, the focal curricula are implicitly compared to a mix of non-green and green curricula.[14]

Table III.4. Timing of schools' curriculum switch, 2019–2020 to 2021–2022

Switch year	Number of schools that switched to Eureka Math	Number of schools that switched to Into Math	Number of schools that did not switch by 2021–2022
2019–2020	19	0	N/A
2020–2021	14	14	N/A
2021–2022	3	2	N/A
Total	36	16	4

Source: School district data.

Additional analyses

In addition to the main difference-in-difference analyses used to estimate the overall effect of schools switching to Eureka Math or Into Math on students' math achievement, we performed four additional types of analyses to help us better interpret the main findings.

Subgroup analyses. To better understand the efficacy of the focal curricula, we explored whether the effects of switching to Eureka Math or Into Math were different for groups of students who are

---

[13] Before switching to Eureka Math or Into Math, 28 schools had been using SpringBoard Mathematics, four schools had been using enVision Math, and one school had been using Ready Mathematics.

[14] Note that information on curriculum use was only available starting in 2019–2020, so we do not know which curricula schools used in the earlier years of the study.

underrepresented in science, technology, engineering, or math (STEM) careers.[15] Specifically, we explored whether the effects of switching to Eureka Math or Into Math differed for students who (1) are Black or Hispanic, or (2) are female.

Sensitivity analyses. Given the complexities of a staggered difference-in-difference analysis, we examined the extent to which results depend on the timing of the curriculum switch and the composition of the comparison schools included. In one sensitivity analysis, we test for evidence that effects differ by the timing of the curriculum switch. In another related analysis, we limit the sample to only schools that switched curricula in 2020–2021 and schools that never switched curricula, so that the analysis is based on a consistent set of baseline and outcome years. Another set of sensitivity analyses limits the comparison group to (1) schools that never switched curricula or (2) schools that eventually switched to the same curriculum.

Bayesian analyses. We used a Bayesian interpretation to characterize the uncertainty of the estimated effects of switching to Eureka Math or Into Math (Deke et al., 2022). The estimated effects from the difference-in-difference analysis can differ from the true effects for several reasons, including error in the measures of math achievement and variability in the sample. For example, only four schools in the study decided not to switch curricula during the study period, but these schools could be poor representatives of the average hypothetical outcomes if all schools had chosen not to switch curricula. Using a Bayesian interpretation enables us to borrow information from other studies to try to reduce this uncertainty and estimate the probability that the effects were greater than different thresholds of interest.

We use the Bayesian Interpretation of Estimates (BASIE) tool from the Institute of Education Sciences to convert the main estimates into probabilities that switching to Eureka Math or Into Math had a positive effect. BASIE uses findings from previously completed education studies to create a prior distribution of effects that describes "how common it is for education interventions to have true (not estimated) effects of varying sizes." In the absence of more information, this prior distribution is the best guess as to whether a given education intervention is likely to be effective. The BASIE tool enabled the study team to obtain an updated posterior distribution that describes how likely it is that switching curricula in the study's context led to positive effects, given what is observed in this study's data and what previous studies examining middle school math outcomes have found.

Descriptive analyses. To better understand contextual factors in the school district, as well as the student and teacher classroom experience, the study team used descriptive methods to analyze survey data collected for the AMS project. As noted earlier, these results draw on data in only 10 schools and are not representative of the entire district. Even so, they offer insight into how people experienced the curricula and how beliefs surrounding math related to students' math achievement. First, we explored the extent to which student and teacher beliefs were correlated with students' math achievement. Second, we tested for differences in the average beliefs of students and teachers by their schools' curriculum choice. Table III.5 summarizes the student and teacher belief outcomes used in these analyses.

---

[15] Although we were interested in whether effects differed for students who are economically disadvantaged, individual-level administrative data on economic disadvantage were not available.

Table III.5. Measures of beliefs used in the descriptive analyses

Construct	Description
Student beliefs	
Student engagement	Positive and active participation in math class, including the desire to meet academic expectations, comply with social and behavioral classroom norms, engage cognitively, and engage emotionally
Math enjoyment	The belief that doing math and being in math class is fun
Math self-efficacy	Confidence in solving math problems and performing math-related tasks
Achievement identity	Identifying and holding a self-concept as someone who can achieve academically; this student belief can improve with intervention or is a strong predictor of future math achievement
Growth mindset	Students' belief that their ability to learn is not fixed but can be developed over time; this is a mindset that can be nurtured in instructional settings
Teacher beliefs	
Perceptions of curriculum and professional learning	
Curriculum consistency	Teacher perceptions of the coherence or alignment of a curriculum with content standards, assessments, and instructional policy (curriculum consistency)
Curriculum specificity	Teacher perceptions of how the level and clarity of detail a curriculum developer provides on instructional content, resources, assessment tasks, and learning objectives (curriculum specificity) influence adherence to an intended curriculum
Curriculum normative authority	The extent to which teachers feel the curriculum is appropriate for their students
Professional learning coherence	The extent to which teachers feel that professional learning content, goals, and activities are consistent with the school curriculum and goals; teacher knowledge and beliefs; the needs of students; and school, district, and state and policies
Professional learning usefulness	The extent to which teachers report the professional learning they participated in is useful to their instruction
Beliefs about school environment	
Collective efficacy	The extent to which teachers believe that they can improve their teaching to positively influence student learning
Collective participation	The extent to which teachers believe their professional learning involves opportunities to build an interactive learning community with other teachers (in the same grade, content area, or school)
Supportive leadership	The extent to which teachers feel encouraged by school leadership to implement learned knowledge and competencies in the classroom
Beliefs about own instruction	
Confidence in culturally relevant mathematics teaching	The extent to which teachers feel confident incorporating culturally responsive math teaching practices into their teaching
Culturally responsive pedagogy	The extent to which teacher beliefs align with teaching practices that recognize that cultural and linguistic differences should be treated as assets for teaching and learning
Confidence meeting student needs (self-efficacy)	The extent to which teachers feel confident teaching students at different achievement levels, of diverse backgrounds, and with individualized learning or language learning needs
Self-reported used of instructional practices	
Ambitious instruction	The frequency with which teachers employ ambitious instructional strategies requiring students to explore multiple representations of mathematics, evaluate

Construct	Description
	representations and approaches to mathematical concepts or procedures, and engage critical thinking skills
Procedural instruction	The frequency with which teachers employ teaching practices that are not conceptually sophisticated, with rote student interactions that do not require them to engage critical thinking skills
Culturally responsive mathematics teaching	The frequency with which teachers incorporate culturally responsive math teaching practices into their teaching

Limitations

The analysis seeks to estimate the causal effect of switching to Eureka Math or Into Math for schools in the district using a quasi-experimental design. However, estimates from difference-in-difference analyses may be biased if trends in student outcomes *before* schools switched curricula were already diverging between schools in the treatment and comparison groups. Figure III.1 shows average outcomes over time for each of our primary outcomes, by schools' curriculum choice. The year just before the switch is indicated by a dashed vertical line. In general, the trends before the switch year are fairly flat, with no obvious pre-trend.

When treatment is staggered, another requirement of the difference-in-difference strategy is that the timing of the treatment is not associated with differential trends in the outcomes. Figure III.2 shows average NWEA MAP z-scores over time, by schools' curriculum choice, separately for each switch year (2019–2020, 2020–2021 and 2021–2022). The lack of notable pre-trends both overall and when comparing groups of schools by adoption year suggests that a difference-in-difference strategy is a reasonable approach for this study, but we cannot rule out the possibility of bias in the results.[16]

Analyses that examine the effects of switching curricula on state test scores suffer from missing data in 2019–2020. This data limitation means that trends in students' state test scores before the switch cannot be assessed effectively for schools that switched after 2019-20, and that schools that switched in 2019-20 cannot contribute to the estimates of the effects after one year. Therefore, the estimated effects of switching curricula on state test scores should be viewed as suggestive evidence and interpreted with caution.

Of note, the curriculum changes examined in this study coincided with the COVID-19 pandemic. The pandemic presented enormous challenges to school administrators and staff, led to drops in student attendance and learning loss, and affected the physical, social, and emotional health of students and their families. Since the pandemic affected all students and schools in the study at the same time, however, it is

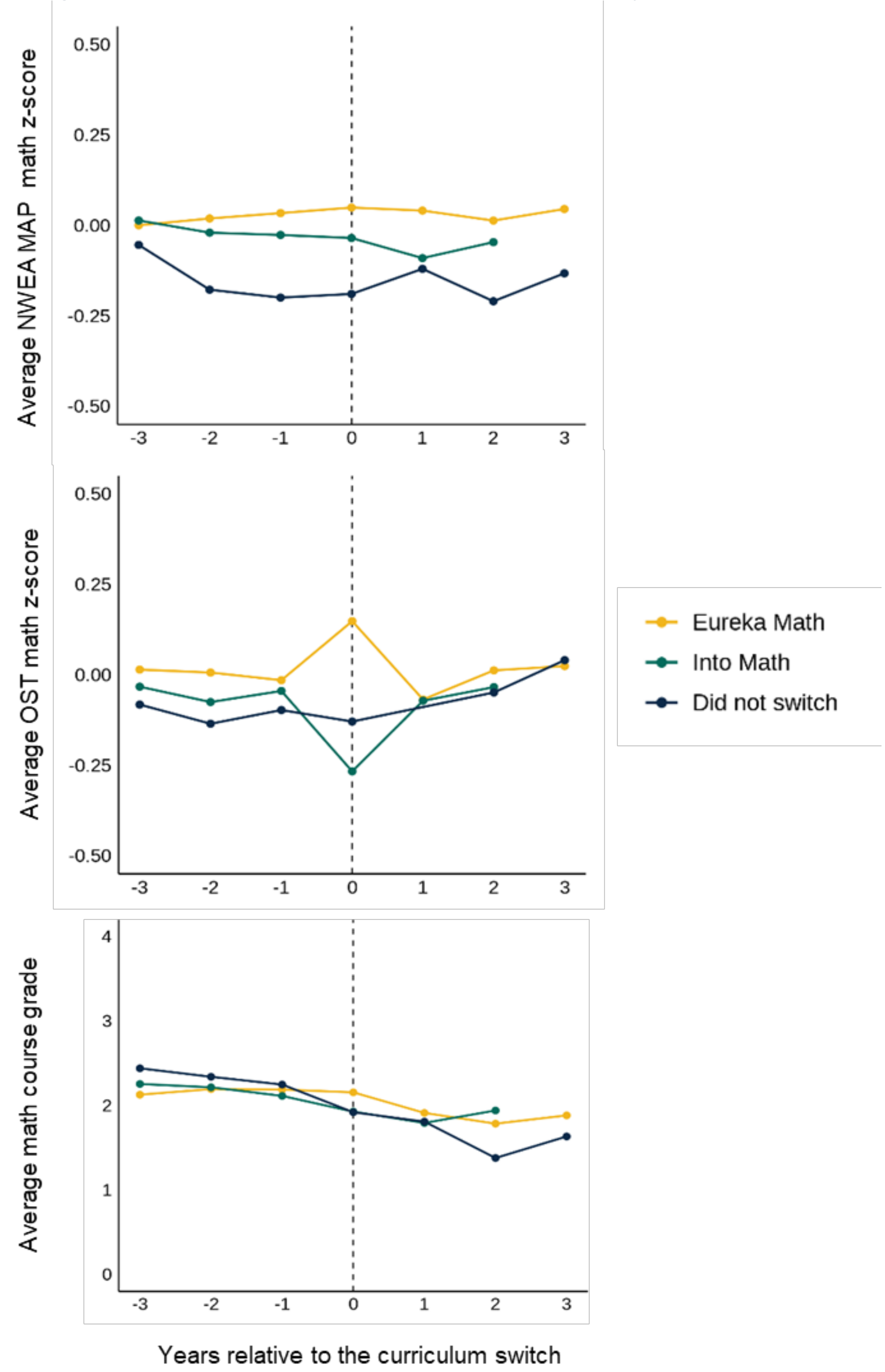
---

[16] Because we are comparing average outcomes over time, the results could also be biased if there are changes in the composition of students over time that differ across curriculum groups. However, as with math achievement outcomes, we find stable trends when examining average student characteristics by schools' curriculum choice over time. Results could also be biased if there are anticipatory effects in advance of the curriculum switch occurring. For example, if teachers expected the upcoming curriculum switch and changed their behavior in advance (for instance, piloting the new curricula in advance), the results could also be capturing the effect of these behavior changes. However, we have no evidence from interviews with school leaders that this happened. Similarly, other policy changes that occurred at the same time in the same schools as the curriculum switches would potentially bias the results. Although we cannot rule out that there might have been other policy changes that coincided with the years in which schools switched curricula, there is no direct evidence that this happened.

unlikely that it would lead to bias in the results. To create bias, it would have to be the case that the schools that switched to Eureka Math or Into Math were also different in their preexisting ability to respond to the pandemic in such a way that caused differences in math achievement during this period. Even so, the difficult context during the pandemic made implementing a new curriculum more challenging and potentially limited the amount and quality of professional development and support teachers received. The results therefore represent the effects of switching a school's curriculum during the unique context of the pandemic. It is possible that, under more normal public health and implementation conditions, the effects of adopting these curricula could be different.

Finally, we note that the analyses of student and teacher survey data are purely descriptive and correspond to a sample of schools purposefully selected to participate in the AMS project. In some cases, they represent only a handful of teachers in each school and should be viewed as case studies. Findings from these analyses should not be interpreted as causal evidence of the effects of adopting the new curricula and should not be generalized to all schools in the district.

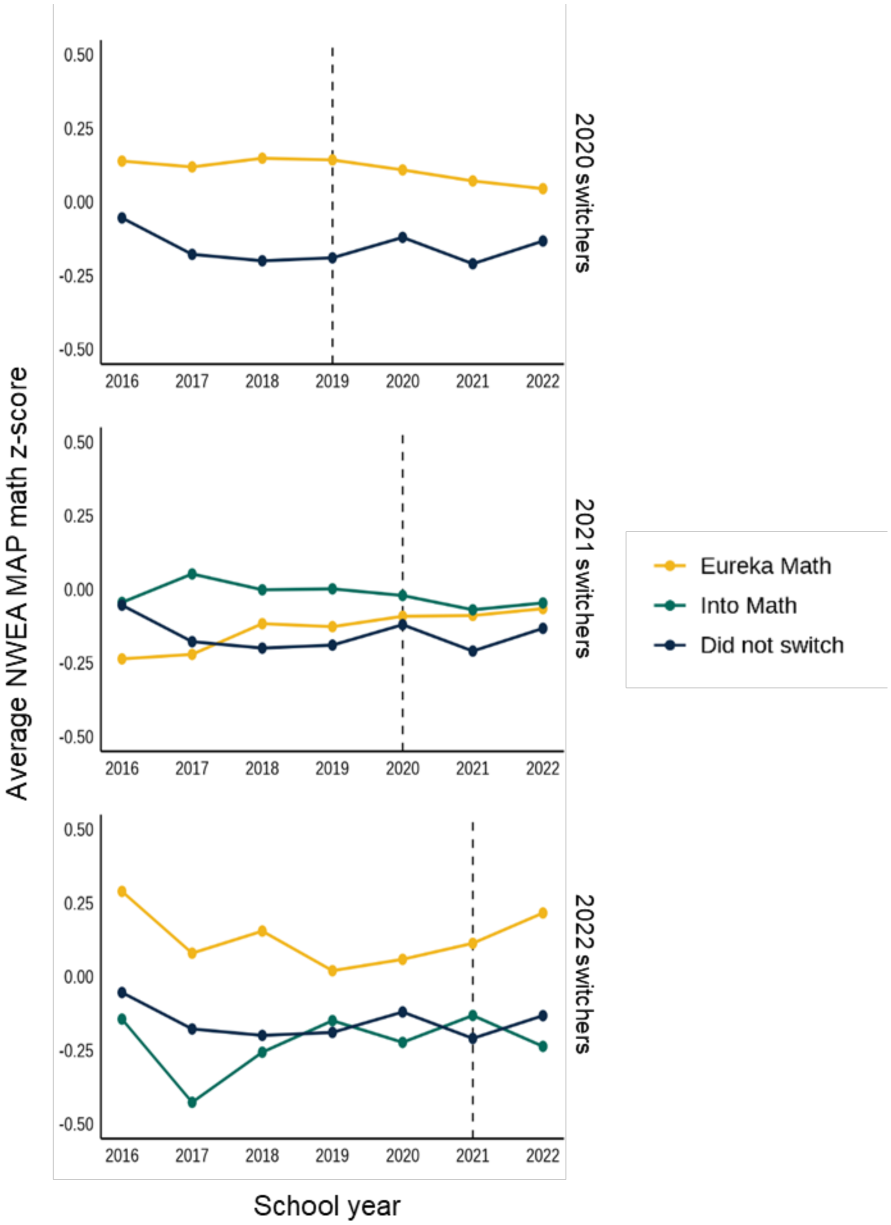
Figure III.1. Average math achievement outcomes over time, by curriculum



Source: District data.

Notes: The data were re-centered so that all years are relative to the timing of the curriculum switch. For example, Year 0 refers to the year before a school switched curricula, Year 1 is the first implementation year, and so on. For schools that did not switch curricula, Year 0 refers to the 2018–2019 school year (the year before schools in the district began adopting new curricula). The year just before the curriculum switch is indicated by a dashed vertical line. The state test was cancelled in the 2019–2020 school year due to the COVID-19 pandemic; therefore, schools that switched curricula in 2019–2020 are missing outcome data from Year 1, schools that switched curricula in 2020–2021 are missing baseline data from Year 0, and so on. This results in incomplete data for the trends shown in these years.

Figure III.2. Average NWEA MAP math z-scores over time, by curriculum and switch year



Source: District data.

Notes: Each year shown indicates the spring semester of the corresponding school year. For example, 2020 refers to the 2019–2020 school year. The year just before the curriculum switch is indicated by a dashed vertical line.

IV. Findings

In this chapter, we present findings from analyses assessing the effects of adopting Eureka Math or Into Math on student math achievement, as well as additional exploratory analyses to support interpretation. First, we present the main efficacy findings based on all schools in the district, a Bayesian interpretation of the main findings, and exploratory results of the effects of the curricula for different student subgroups. Then, we present descriptive findings from student and teacher survey data from a subset of schools, including differences in the beliefs of students and teachers in schools that used Eureka Math versus Into Math.

Main efficacy findings

In what follows, we summarize the results of the main difference-in-difference analyses, which estimate the effects on students of switching curricula in the average school after each year of implementation. We focus on estimates that are statistically significant at the 5 percent level or better. Results from additional sensitivity analyses appear in Appendix B.

Switching to Eureka Math or Into Math led to increases in students’ math course grades but had no detectable effect on their achievement on standardized math tests.

Table IV.1 presents the estimated effects of switching to either focal curriculum after one, two, and three years of implementation, without distinguishing between Eureka Math and Into Math.[17] We find no evidence of a statistically significant impact on either NWEA MAP or state test math standardized test scores or the percentage of students who meet proficiency thresholds, on average. In contrast, switching to a focal curriculum is estimated to increase students’ math GPA by 0.37 points on a 4.0 scale after one year, 0.36 points after two years, and 0.33 points after three years. These estimates are equivalent to an increase of about 0.3 standard deviations relative to the districtwide middle school math GPA distribution in the 2018–2019 school year (the last year before schools began to switch curricula). Likewise, switching curricula led to the percentage of students with a passing GPA (that is, above 2.0) increasing by 12 percentage points after one, two, and three years. These findings suggest a consistent rise in students’ average math GPA as well as the percentage of students with passing grades. However, we cannot distinguish whether this results from a change in students’ math knowledge or skills, or from schools that adopted new curricula also adopting new grading standards.

---

[17] Because all schools that switched curricula in 2019–2020 chose Eureka Math, the three-year results represent the effects of switching to Eureka Math only.

Table IV.1. Effects of switching to a focal curriculum

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
After one year	0.01 (0.04)	0.00 (0.01)	0.06 (0.15)	0.04 (0.04)	0.37** (0.07)	0.12** (0.04)
After two years	-0.02 (0.05)	0.00 (0.00)	0.01 (0.15)	-0.01 (0.04)	0.36** (0.07)	0.12** (0.04)
After three years	-0.08 (0.07)	-0.01 (0.01)	-0.12 (0.17)	-0.07 (0.05)	0.33* (0.13)	0.12* (0.06)
N students	19,704	19,704	18,775	18,775	19,642	19,642
N schools	56	56	56	56	56	56
N years	7	7	6	6	7	7

Source: School District data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. The effects after three years only reflect the effect of switching to Eureka Math. Standard errors are shown in parentheses.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

There were few detectable differences in the effects of switching to Eureka Math compared to Into Math.

Table IV.2 presents the estimated effects of switching to Eureka Math and Into Math separately. As with the overall effects of switching to either focal curriculum, we find no evidence of a statistically significant effect of switching to either Eureka Math or Into Math on students' math test scores or proficiency. However, both curricula led to similar increases in students' math course grades. Switching to Eureka Math is estimated to increase students' math GPA by between 0.39 points after one year of implementation (or about 0.4 standard deviations) and 0.33 points after three years (or about 0.3 standard deviations). It is also estimated to significantly increase the percentage of students with passing math grades by 13 percentage points after one year and 12 percentage points after two and three years. Similarly, switching to Into Math is estimated to increase students' math GPA by 0.33 points (about 0.3 standard deviations) after both one and two years. It also significantly increased the percentage of students with passing math grades by 11 percentage points after one year and 10 percentage points after two years.

For most outcomes, there is no detectable difference between the effect of switching to Eureka Math and switching to Into Math. Neither curriculum had a statistically significant effect on math test scores or proficiency, and both had similarly sized, positive effects on math grades and passing rates. There is one exception: There is some evidence that the effect on NWEA MAP scores after one year of implementation may be larger for schools that chose Eureka Math compared to Into Math. The difference between these two estimates is statistically significant, though neither can be distinguished from zero.

Table IV.2. Effects of switching to a focal curriculum, separately for Eureka Math and Into Math

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Eureka Math after one year	0.04† (0.04)	0.00 (0.01)	0.06 (0.15)	0.05 (0.04)	0.39** (0.08)	0.13** (0.04)
Into Math after one year	-0.04† (0.05)	0.00 (0.01)	0.08 (0.16)	0.03 (0.04)	0.33** (0.11)	0.11* (0.05)
Eureka Math after two years	-0.01 (0.05)	0.01 (0.00)	0.02 (0.15)	-0.01 (0.04)	0.35** (0.08)	0.12** (0.04)
Into Math after two years	-0.09 (0.06)	-0.01 (0.01)	-0.04 (0.15)	-0.01 (0.03)	0.33** (0.10)	0.10* (0.05)
Eureka Math after three years	-0.10 (0.07)	-0.01 (0.01)	-0.13 (0.17)	-0.07 (0.05)	0.33* (0.12)	0.12* (0.06)
N students	19,704	19,704	18,775	18,775	19,642	19,642
N schools	56	56	56	56	56	56
N years	7	7	6	6	7	7

Source: School district data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses.

† Difference between Eureka Math and Into Math after the same number of years is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Bayesian interpretation of the findings

The results from the difference-in-difference analyses are estimates of the true effects that are measured with uncertainty, meaning it can be difficult to detect statistically significant impacts due to reasons such as error in the measures of math achievement and variability in the sample. Below, we use a Bayesian interpretation to characterize the uncertainty of the results (Deke et al., 2022). Using a Bayesian interpretation enables us to borrow information from other studies to try to reduce this uncertainty and estimate the probability that the curricula led to meaningful positive effects.[18]

Switching to Eureka Math likely led to positive effects on standardized test scores in the first year, but this probability diminished after each year of implementation.

Using BASIE to estimate the probability that switching to Eureka Math or Into Math had a meaningful positive effect, we find two main patterns (Table IV.3). First, the probability of positive effects for both curricula decreased with time after the curricula were adopted. For example, the effect of switching to

---

[18] The resulting probability statements are retrospective statements about the effect of switching curricula in the context studied and not predictions of what would happen in other contexts. In addition, probability statements made using BASIE are relative to the specific prior distribution used—in our case, this prior distribution is based on effects from other educational evaluations examining middle school math achievement that have been reviewed by the What Works Clearinghouse and met its standards of rigor. Using a different set of studies could lead to different results.

Eureka Math on NWEA MAP math scores was very likely positive after one year (81 percent probability of an effect greater than zero) but got closer to even odds after two years (44 percent) and was most likely negative after three years (11 percent). The probability that switching to Into Math had a positive effect on NWEA MAP scores also decreased over time, though this likelihood was low even in the first year (25 percent). The pattern of the likelihood of positive effects decreasing over time also holds for average state test scores and proficiency rates on both the NWEA MAP and the state test. Given that proficiency thresholds are higher for the NWEA MAP than the state test, it is perhaps unsurprising that the probability of positive effects of switching curricula on NWEA MAP proficiency is at best slightly greater than even, compared to a maximum 86 percent chance of positive effects of switching to Eureka Math and a 71 percent chance of positive effects of switching to Into Math on the state test proficiency after one year.

Switching to either Eureka Math or Into Math very likely led to meaningful improvements in math GPA.

The other major pattern evident from applying a Bayesian interpretation to the results is the very high likelihood of positive effects on math course grades. We can be very sure of positive effects on grades for both curricula; the estimated probabilities are near 100 percent, as shown in Table IV.3. One benefit of using BASIE is the ability to report probabilities that effects are greater than a threshold that may be deemed meaningful. Even when assessing the likelihood of effects larger than zero—for example, the likelihood that switching curricula increased average math GPA by *at least* 0.2 points on a 4.0 scale (equivalent to about 0.2 standard deviations, which is typically considered the threshold for a large effect size)—we find probabilities that exceed 70 percent for both curricula and all years of implementation (results not shown).

Table IV.3. Bayesian interpretation of the main efficacy findings

	NWEA MAP math				State test math				Math courses			
	Average z-score		Proficiency rate		Average z-score		Proficiency rate		Average math GPA		Passing rate	
	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0
Eureka Math after one year	0.04† (0.04)	81%	0.00 (0.01)	53%	0.06 (0.15)	73%	0.05 (0.04)	86%	0.39** (0.08)	100%	0.13** (0.04)	100%
Into Math after one years	-0.04† (0.05)	25%	0.00 (0.01)	52%	0.08 (0.16)	75%	0.03 (0.04)	71%	0.33** (0.11)	100%	0.11* (0.05)	98%
Eureka Math after two years	-0.01 (0.05)	44%	0.01 (0.00)	58%	0.02 (0.15)	66%	-0.01 (0.04)	41%	0.35** (0.08)	100%	0.12** (0.04)	100%
Into Math after two years	-0.09 (0.06)	12%	-0.01 (0.01)	40%	-0.04 (0.15)	54%	-0.01 (0.03)	38%	0.33** (0.10)	100%	0.10* (0.05)	97%
Eureka Math after three years	-0.10 (0.07)	11%	-0.01 (0.01)	34%	-0.13 (0.17)	41%	-0.07 (0.05)	9%	0.33* (0.12)	99%	0.12* (0.06)	97%

Source: School District data.

Notes: The estimated effects shown are the same as those in Table IV.2. Standard errors are shown in parentheses. The probabilities shown reflect the likelihood that these effects were greater than the noted thresholds and were calculated using BASIE. Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level.

† Difference between Eureka Math and Into Math after the same number of years is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Exploratory efficacy findings by student subgroup

To understand whether switching to Eureka Math or Into Math was especially effective for different student subgroups of interest, we explored whether there were differential effects for students who are Black, Hispanic, or female.[19] Tables IV.4 and IV.5 present the results of these exploratory analyses. The top panel in each table shows the estimated differences in the effects of the curricula between the subgroup of interest and the reference group. For example, the top panel of Table IV.5 shows the estimated differences in the effects of the curricula for female versus male students, where a positive (negative) coefficient indicates that the effect was larger (smaller) for female than male students. The bottom panel shows the estimated effects for the reference group—for example, male students. Descriptive comparisons of math achievement and belief outcomes by student subgroup are available in Appendix B.

Both focal curricula were relatively less effective in raising the performance of Black and Hispanic students on the state standardized test compared to other students.

Table IV.4 presents the results for Black and Hispanic students compared to students of other races and ethnicities. [20] The results suggest a few differential effects by race and ethnicity groups, which are summarized below for each focal curriculum:

Eureka Math

- Switching to Eureka Math was less effective in raising the state test math proficiency rates of Black and Hispanic students compared to students of other races or ethnicities. After one year of implementation, Eureka Math led to an increase in the percentage of students of other races and ethnicities proficient on the state test by 19 percentage points. In contrast, the effect for Black and Hispanic students was 15 percentage points lower. Although the positive effect on the state test proficiency for students of other races and ethnicities fades over time, the effect for Black or Hispanic students remains lower in comparison.
- Switching to Eureka Math was more effective in raising the percentage of Black and Hispanic students with passing math grades, at least in the second year of implementation. After two years of implementation, this effect was 16 percentage points higher for Black and Hispanic students compared to students of other races or ethnicities (for whom the effects are small and not statistically significant). The differences in effects between groups are not statistically significant in the first or third years of implementation.

---

[19] Because all students in the school district are eligible for free or reduced-price meals and no other measures of socioeconomic disadvantage were available, we could not measure how the effects differed for students experiencing poverty.

[20] In the school district, 64 percent of students identified as Black in the 2018–2019 school year, 17 percent identified as Hispanic, 15 percent identified as White, and 4 percent identified as other races or ethnicities. Given the relatively smaller population of Hispanic students, as well as the similarity of results for Black and Hispanic students, we present findings for Black and Hispanic students as one group. Results are reported for Black and Hispanic students separately in Appendix B.

- There is no evidence that switching to Eureka Math had differential effects on NWEA MAP scores, NWEA MAP proficiency rates, state test scores, or math GPA for Black and Hispanic students compared to students of other races or ethnicities in any year of implementation.

Table IV.4. Effects of switching to Eureka Math and Into Math, by students' race and ethnicity

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	**Proficient (%)**	**z-score (mean)**	**Proficient (%)**	**GPA (mean)**	**Passing GPA (%)**
Difference in effects for Black and Hispanic students compared to other races or ethnicities						
Eureka Math after one year	0.04	-0.05	-0.11†	-0.15*	0.18	0.08
	(0.10)	(0.08)	(0.16)	(0.07)	(0.15)	(0.06)
Into Math after one year	-0.02	-0.03	-0.32*†	-0.12	0.20	0.10
	(0.11)	(0.08)	(0.16)	(0.07)	(0.16)	(0.08)
Eureka Math after two years	-0.01	-0.05	-0.08	-0.10	0.26	0.16*
	(0.10)	(0.08)	(0.15)	(0.06)	(0.16)	(0.06)
Into Math after two years	-0.08	0.01	-0.10	-0.12	0.11	0.06
	(0.11)	(0.09)	(0.16)	(0.06)	(0.17)	(0.07)
Eureka Math after three years	0.02	-0.04	-0.03	-0.16*	0.14	0.11
	(0.11)	(0.09)	(0.16)	(0.07)	(0.16)	(0.06)
Effects for students of other races or ethnicities (reference group)						
Eureka Math after one year	-0.00	0.05	0.18†	0.19*	0.23	0.05
	(0.07)	(0.08)	(0.22)	(0.09)	(0.13)	(0.06)
Into Math after one year	-0.02	0.03	0.37†	0.14	0.14	0.02
	(0.08)	(0.08)	(0.21)	(0.09)	(0.16)	(0.08)
Eureka Math after two years	-0.01	0.05	0.11	0.09	0.12	-0.02
	(0.07)	(0.08)	(0.21)	(0.08)	(0.15)	(0.06)
Into Math after two years	-0.03	-0.01	0.07	0.11	0.22	0.04
	(0.08)	(0.09)	(0.22)	(0.07)	(0.16)	(0.07)
Eureka Math after three years	-0.13	0.03	-0.07	0.09	0.19	0.01
	(0.08)	(0.08)	(0.22)	(0.09)	(0.20)	(0.08)
N students	19,704	19,704	18,775	18,775	19,642	19,642
N schools	56	56	56	56	56	56
N years	7	7	6	6	7	7

Source: School district data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses. The top panel shows the estimated differences in the effects of the curricula between (1) the subgroup of interest and (2) the corresponding reference group; a positive (negative) coefficient there indicates the effect was larger (smaller) for the subgroup of interest compared to the reference group. The bottom panel shows the estimated effects for the reference group.

† Difference for students of the same race or ethnicity between Eureka Math and Into Math after the same number of years is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Into Math

- Switching to Into Math was less effective in raising the state test math scores of Black and Hispanic students compared to students of other races or ethnicities, at least in the first year of implementation. On the state test math test, the effect for Black and Hispanic students is 0.3 standard deviations lower than for students of other races or ethnicities. The difference in effects between groups is not statistically significant in the second year.
- There is no evidence that switching to Into Math had differential effects on NWEA MAP scores, NWEA MAP proficiency rates, state test proficiency rates, math GPA, or math passing rates for Black and Hispanic students compared to students of other races or ethnicities in any year of implementation.

Eureka Math was relatively less effective in raising the performance of female students on standardized tests compared to male students, and relatively more effective in raising their math course grades. In contrast, Into Math had similar effects for both male and female students but led only to increases in students' math course grades.

Table IV.5 presents the results for female students compared to male students. The results suggest some differential impacts by gender for Eureka Math but not Into Math, as summarized below:

Eureka Math

- Switching to Eureka Math was less effective in raising the math performance of female students compared to male students on some test outcomes. For example, switching to Eureka Math significantly raised the percentage of male students scoring proficient on the NWEA MAP by 2 percentage points after one and two years. Meanwhile, this effect was 3 to 4 percentage points lower for female students in each year. This pattern is echoed for average NWEA MAP scores in Year 3 and for average state test math scores in all years. The effects on state test proficiency are not statistically different by gender.
- Switching to Eureka Math was more effective in raising math course grades for female students compared to male students. For male students, switching to Eureka Math led to an increase in math GPA of 0.28 points on a 4-point scale after one year, 0.3 points after two years, and 0.23 points after three years (all but the last are statistically significant). For female students, the effect was even larger by about 0.2 points, with the difference being statistically significant after one and three years. There is also some evidence that the effect on passing rates was different for female students, though only one year after switching to Eureka Math, when the effect was 9 percentage points higher for female students.

Into Math

- Switching to Into Math had similar effects for both female and male students, leading to increases in their math course grades but no other outcomes. After one and two years of implementation, both male and female students experienced significant increases in their math GPA (of about 0.3 points in each year) and their probability of receiving a passing grade (of 10 to 11 percentage points). We did not detect any differences in effects by gender.

Table IV.5. Effects of switching to Eureka Math and Into Math, by students' gender

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Difference in effects for female students compared to male students						
Eureka Math after one year	-0.04	-0.04**†	-0.11*	-0.00	0.24*	0.09*
	(0.04)	(0.01)	(0.05)	(0.03)	(0.09)	(0.04)
Into Math after one year	0.02	-0.01†	-0.03	-0.03	0.17	0.03
	(0.05)	(0.01)	(0.05)	(0.02)	(0.09)	(0.04)
Eureka Math after two years	-0.05	-0.03*	-0.15**†	-0.04	0.12	0.07
	(0.05)	(0.01)	(0.04)	(0.02)	(0.08)	(0.04)
Into Math after two years	0.02	-0.01	-0.02†	0.00	0.05	-0.01
	(0.06)	(0.01)	(0.04)	(0.02)	(0.09)	(0.05)
Eureka Math after three years	-0.09*	-0.03**	-0.11**	-0.02	0.20*	0.08
	(0.04)	(0.01)	(0.04)	(0.03)	(0.09)	(0.04)
Effects for male students (reference group)						
Eureka Math after one year	0.06†	0.02*	0.12	0.05	0.28**	0.09*
	(0.05)	(0.01)	(0.14)	(0.04)	(0.09)	(0.04)
Into Math after one year	-0.05†	0.01	0.09	0.04	0.24*	0.10*
	(0.06)	(0.01)	(0.15)	(0.04)	(0.12)	(0.05)
Eureka Math after two years	0.01†	0.02*	0.09†	0.01	0.30**	0.09*
	(0.06)	(0.01)	(0.15)	(0.04)	(0.08)	(0.04)
Into Math after two years	-0.10†	-0.00	-0.03†	-0.01	0.31**	0.11*
	(0.08)	(0.01)	(0.14)	(0.03)	(0.11)	(0.04)
Eureka Math after three years	-0.06	0.01	-0.08	-0.06	0.23	0.08
	(0.08)	(0.01)	(0.16)	(0.05)	(0.14)	(0.06)
N students	19,704	19,704	18,775	18,775	19,642	19,642
N schools	56	56	56	56	56	56
N years	7	7	6	6	7	7

Source: School District data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses. The top panel shows the estimated differences in the effects of the curricula between (1) the subgroup of interest and (2) the corresponding reference group; a positive (negative) coefficient there indicates the effect was larger (smaller) for the subgroup of interest compared to the reference group. The bottom panel shows the estimated effects for the reference group.

† Difference for students of the same gender between Eureka Math and Into Math after the same number of years is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Exploratory survey findings

To support the interpretation of the main efficacy findings, we analyzed data from surveys that measured the perceptions of students and teachers in 10 schools that adopted Eureka Math or Into Math. First, we explored the extent to which these student and teacher beliefs were correlated with students' math test scores and course grades. Then, we assessed whether, on average, there were differences in the beliefs of students and teachers based on their school's curriculum. Additional details about each belief scale, including its definition, survey items, and value ranges, appear in Tables A.6 and A.7 in Appendix A.

Students' beliefs about their achievement identity and math self-efficacy had the strongest positive relationships to their math achievement.

In general, the student belief constructs measured in this study are positively—though weakly—correlated with math achievement, as measured by students' scores on the NWEA MAP and state math tests and their math course GPA (Table IV.7). Consistent with other research literature, students' achievement identity and math self-efficacy had the strongest relationships to these achievement outcomes, with correlation coefficients ranging from 0.3 to 0.4. To put these correlations into context, consider that students with a high achievement identity (or high math self-efficacy) are about twice as likely as students with a low achievement identity (or low math self-efficacy) to score at or above proficiency on the state test math test.

Table IV.7. Correlations between student beliefs and their math achievement

Construct	NWEA math z-score	State test math z-score	Math course GPA
Achievement identity	0.33	0.36	0.42
Growth mindset	0.22	0.15	0.11
Math enjoyment	0.09	0.18	0.15
Math self-efficacy	0.37	0.33	0.31
Student engagement	0.22	0.25	0.20

Source: Student survey data.

Notes: This table shows the pairwise Pearson correlation coefficients between student math belief constructs and student math achievement outcomes. Correlation coefficients can range from -1 to 1, with higher absolute values indicating a stronger relationship and positive (negative) values indicating a positive (negative) relationship. The sample sizes for these correlations range between 266 and 302 students.

To understand how students' math experiences differ between curricula, we compared the average scores of students in Eureka Math versus Into Math schools. Because each curriculum group of schools has different characteristics (for example, 26 percent of students in Eureka Math schools have special education needs, compared to 19 percent of students in Into Math schools), we also compared average beliefs after (1) controlling for differences in the characteristics of students and their schools and (2) controlling for students' beliefs at the start of the school year. Table IV.8 presents the findings.

Otherwise similar students had less positive beliefs around math when taught using Eureka Math than Into Math.

After controlling for differences in student and school characteristics, we find that students in Eureka Math schools had lower scores, on average, on achievement identity, math enjoyment, math self-efficacy, and student engagement than students in Into Math schools. For example, students in Eureka Math schools scored about one point lower, on average, on the achievement identity and math self-efficacy scales (which range from 1 to 6) compared to their peers in Into Math schools.

Although we cannot attribute these differences to the curricula, the results suggest that, on average, otherwise similar students had less positive experiences when taught using Eureka Math than Into Math. The differences between students in Eureka Math and Into Math hold true even when also accounting for students' beliefs at the start of the school year (Table IV.8). In other words, students in Eureka Math schools experienced less growth in these beliefs than students in Into Math schools between the fall and spring semesters. In fact, we also find that in schools that used Eureka Math, students' average math enjoyment *decreased* over the course of the year.[21] In contrast, there were no statistically significant changes in student beliefs between fall and spring among students in schools using Into Math.

Table IV.8. Differences in student beliefs between Eureka Math and Into Math Schools

Construct	Raw difference between Eureka Math and Into Math schools		Difference after accounting for student and school characteristics		Difference after also accounting for baseline student beliefs in fall	
	Estimated difference	Number of students	Estimated difference	Number of students	Estimated difference	Number of students
Achievement identity	-0.03 (0.23)	302	-1.32** (0.13)	211	-1.03** (0.04)	188
Growth mindset	0.01 (0.16)	298	0.18 (0.15)	209	0.02 (0.14)	186
Math enjoyment	-0.18 (0.09)	298	-0.85** (0.17)	209	-0.68** (0.10)	185
Math self-efficacy	-0.12 (0.20)	295	-0.67** (0.14)	207	-1.11** (0.10)	184
Student engagement	-0.07 (0.13)	300	-0.48* (0.15)	210	-0.29* (0.10)	187

Source: Student survey data.

Notes: This table shows the difference in average beliefs between surveyed students in Eureka Math and Into Math schools, reported in the units of each survey scale (see Appendix Table A.6). Standard errors are shown in parentheses. For each construct, the first set of columns shows the raw difference in means, and the second set of columns shows the estimated difference from an ordinary least squares regression that controls for a student's demographics (race/ethnicity, gender, English language proficiency, and special education status), baseline math achievement (NWEA MAP z-score, state test z-score, and math course GPA), attendance, and suspensions; and school fixed effects. The third set of columns additionally controls for the student's baseline beliefs at the start of the school year.

** Difference is statistically significant at the 1 percent level.

* Difference is statistically significant at the 5 percent level.

---

[21] In the fall, students had an average math enjoyment score of 3.5 on a scale ranging from 1 to 5; the same students scored an average of 3.3 in the spring. This decrease is statistically significant. No other belief constructs had statistically significant changes between fall and spring for either curriculum.

Teachers' views of their collective efficacy and their school leadership had the strongest positive relationships to their students' math achievement.

The extent to which the various teacher belief constructs assessed are correlated with the math achievement of teachers' students varies considerably, as shown in Table IV.9. Few constructs consistently exhibit strong correlations (greater than 0.5 in absolute value) with achievement outcomes.

Teachers who report higher levels of collective efficacy—that is, the belief that teachers in their school as a group can improve their teaching to meet the needs of all students—and who report having more supportive school leadership have students with higher math achievement on both standardized tests and math course grades. Collective efficacy in particular appears to be a strong predictor of student achievement, with correlation coefficients of 0.7 or greater across all three achievement outcomes.

Teachers who report using ambitious instructional strategies with greater frequency also have students with higher math course grades, though the correlation between teachers' reported used of ambitious instruction and students' math test scores is relatively weaker. Teachers' reported frequency of using culturally responsive instructional practices was not strongly related to their students' math achievement, nor was their belief in the importance of culturally responsive pedagogy or confidence with culturally responsive math teaching. Interestingly, teachers with greater confidence in their ability to meet their students' needs tended to have students with lower levels of math achievement.

Teachers' perceptions of their curricula are not consistently related to their students' math achievement; however, there is a strong, though negative, correlation between students' math achievement and teachers' views of whether the professional learning they received is coherent with their school curriculum and goals, district policies, and their needs. Teachers' views on the usefulness of their professional learning also showed a negative correlation with their students' math achievement. There may be multiple reasons for the observed relationships. For example, it is possible that teachers in schools where students had greater academic needs participated in more, or better-quality, professional learning.

Table IV.9. Correlations between teacher beliefs and students' math achievement

Construct	NWEA math z-score	State test math z-score	Math course GPA
Perceptions of curriculum and professional learning			
Curriculum consistency	-0.04	-0.06	0.28
Curriculum specificity	-0.15	-0.05	0.12
Curriculum normative authority	-0.27	-0.19	-0.17
Professional learning coherence	-0.53	-0.80	-0.70
Professional learning usefulness	-0.49	-0.44	-0.36
Beliefs about school environment			
Collective efficacy	0.69	0.76	0.77
Collective participation	0.10	0.21	0.22
Supportive leadership	0.52	0.42	0.55

Construct	NWEA math z-score	State test math z-score	Math course GPA
Beliefs about own instruction			
Confidence with culturally responsive mathematics teaching	0.17	0.24	0.20
Belief in culturally responsive pedagogy	-0.08	-0.23	0.00
Confidence meeting student needs (self-efficacy)	-0.49	-0.41	-0.41
Self-reported use of instructional practices			
Ambitious instruction	0.10	0.38	0.56
Procedural instruction	0.13	0.23	0.11
Culturally responsive mathematics teaching	-0.24	-0.02	0.09

Source: Teacher survey data.

Notes: This table shows the pairwise Pearson correlation coefficients between teacher belief constructs and their students' math achievement outcomes. Correlation coefficients can range from -1 to 1, with higher absolute values indicating a stronger relationship and positive (negative) values indicating a positive (negative) relationship. The sample sizes for these correlations range between 9 and 15 teachers.

To understand how teachers' perceptions differed, we compared the average responses of teachers in Eureka Math schools versus Into Math schools on the available scales. Although the characteristics of teachers and students in each curriculum group of schools differs, we were unable to account for these differences due to the small number of teachers surveyed; the number of respondents for each question ranged from nine to 15 teachers in total. Therefore, we focus on comparing raw differences, as shown in Table IV.10.[22] Given the small sample sizes, we are unable to detect whether any differences are statistically different from zero. In addition, these comparisons do not represent all middle school math teachers, either in surveyed schools or across the district, and any differences in respondents' beliefs cannot be attributed to the curricula.

Compared to teachers in Into Math schools, teachers in Eureka Math schools were more likely to report their curriculum was coherent but less likely to feel it was appropriate for their students.

On average, teachers in Eureka Math schools had stronger perceptions of the coherence of the curriculum with content standards, assessments, and instructional policy (curriculum consistency). Teachers in both groups had similar beliefs regarding the specificity of the curricula—that is, the level and clarity of detail provided to implement it. However, teachers in Eureka Math schools were less likely to state the curriculum was appropriate for their students (curriculum normative authority). In terms of the professional learning they received, Eureka Math teachers were less likely to view it as coherent but more likely to find it useful. (These questions asked about all professional learning around math instruction, not just for the curricula.)[23] In effect size units, these differences range from 0.19 to 0.36 standard deviations.

---

[22] Because the teacher construct scales vary widely in the range of possible values (as shown in Appendix Table A.7), we focus the discussion on differences measured using standard deviation units to aid in interpretation.

[23] In a separate analysis (not shown), we found no notable differences in the amount of professional learning teachers reported receiving or how much it specifically focused on the curricula. However, a quarter of both Eureka Math and Into Math teachers who responded to this survey question reported that "none" of the professional learning they received in the previous year specifically focused on the curricula, and about half said only "a little" of it specifically focused on the curricula (N = 12). Nine of the 10 schools surveyed were in the second year of implementation.

Compared to teachers in Into Math schools, teachers in Eureka Math schools were more likely to report collaborative school environments and supportive leadership but reported lower levels of collective self-efficacy.

Teachers in Eureka Math schools had stronger beliefs that collective planning time was used to build an interactive learning community with other teachers (effect size = 0.59) and that they felt supported by their school leadership (effect size = 1.19). However, when asked about their school's collective self-efficacy to meet students' needs, Eureka Math teachers had less positive views than Into Math teachers (effect size = -0.70).

Compared to teachers in Into Math schools, teachers in Eureka Math schools were more likely to believe in culturally responsive pedagogy but were less confident in their ability to meet their students' needs.

On average, teachers in Eureka Math schools had stronger views about the importance of culturally responsive pedagogy—that is, the belief that cultural and linguistic differences should be treated as assets for teaching and learning—compared to teachers in schools using Into Math (effect size = 0.50). Both groups had similar levels of confidence with culturally responsive mathematics teaching. However, Eureka Math teachers had lower levels of confidence in their ability to meet the needs of students at different achievement levels, of diverse backgrounds, and with individualized learning or language learning needs (effect size = -0.49). As noted above, these teachers also reported lower levels of collective self-efficacy.

Compared to teachers in Into Math schools, teachers in Eureka Math schools were more likely to report using ambitious and culturally responsive instruction and less likely to report using procedural instruction.

When asked about the frequency with which they used various types of instructional strategies, teachers in Eureka Math schools were more likely than teachers in Into Math schools to report using ambitious instructional strategies requiring students to explore multiple representations of mathematics, evaluate representations and approaches to mathematical concepts or procedures, and engage critical thinking skills (effect size = 0.42), as well as culturally responsive instructional strategies (effect size = 0.38).[24] In contrast, they were less likely to report using procedural instructional strategies that are not conceptually sophisticated and do not require students to engage critical thinking skills (effect size = -1.01).

---

[24] Data from classroom observations, in which a total of four teachers were observed twice (two from Eureka Math schools and two from Into Math schools), showed the opposite pattern than the self-reported survey data. Based on these classroom observations, Into Math teachers received higher scores on measures of ambitious instruction and CRMT compared to Eureka Math teachers. Observations used the M-Scan tool and a CRMT classroom observation tool that the AMS study team developed. Differences between the observation and survey data could be due to being based on different samples of teachers or the different instruments and respondent types. For example, survey self-reports may be prone to social-desirability bias and may therefore be less reliable than classroom observations.

Table IV.10. Differences in teacher beliefs and practices between Eureka Math and Into Math schools

Construct	Average in Eureka Math schools	Number of teachers	Average in Into Math schools	Number of teachers	Difference in scale units	Difference in effect size units
Perceptions of curriculum and professional learning						
Curriculum consistency	4.27 (0.42)	10	4.10 (0.56)	5	0.17	0.36
Curriculum specificity	3.98 (0.38)	10	3.99 (0.54)	5	-0.00	-0.01
Curriculum normative authority	3.19 (0.51)	10	3.28 (0.50)	5	-0.10	-0.19
Professional learning coherence	4.90 (1.32)	8	5.14 (0.28)	4	-0.24	-0.21
Professional learning usefulness	2.43 (1.32)	8	2.14 (0.51)	4	0.30	0.26
Beliefs about school environment						
Collective efficacy	5.01 (0.95)	8	5.60 (0.51)	4	-0.59	-0.70
Collective participation	2.30 (1.05)	7	1.70 (0.71)	2	0.60	0.59
Supportive leadership	5.09 (1.29)	8	6.44 (0.31)	4	-1.36	-1.19
Beliefs about own instruction						
Confidence with culturally responsive mathematics teaching	7.15 (1.67)	10	7.27 (0.86)	5	-0.12	-0.08
Belief in culturally responsive pedagogy	3.69 (0.51)	10	3.92 (0.36)	5	-0.24	0.50
Confidence meeting student needs (self-efficacy)	2.23 (1.35)	8	2.83 (0.91)	4	-0.60	-0.49
Self-reported use of instructional practices						
Ambitious instruction	2.35 (0.40)	8	2.16 (0.59)	4	0.20	0.42
Procedural instruction	1.48 (0.38)	8	1.95 (0.55)	4	-0.47	-1.08
Culturally responsive mathematics teaching	1.62 (0.57)	8	1.39 (0.71)	4	0.23	0.38

Source: Teacher survey data.

Notes: This table shows the difference in average reported beliefs and practices between surveyed teachers in Eureka Math and Into Math schools, reported in the units of each survey scale (see Appendix Table A.7) as well as effect size units calculated using Cohen's *d*. Standard deviations are shown in parentheses. It was not possible to control for differences in teachers' characteristics or classroom students in this analysis due to the small sample sizes. None of the differences is statistically significant.

V. Conclusion

This chapter concludes the report with a discussion of the findings and plans for future research.

Discussion

Starting in the 2019–2020 school year, the district's leadership provided updated guidance to schools around their math curricula selection. Schools were given two options vetted by the district: Eureka Math and Into Math. Of the 56 middle schools in the efficacy study, 36 schools (64 percent) adopted Eureka Math, 16 schools (29 percent) adopted Into Math, and four schools (7 percent) did not switch curricula by 2021–2022. Most of the schools that adopted a new curriculum switched from SpringBoard Mathematics, a non-green rated curriculum, while the schools that did not switch continued using Ready Mathematics or enVision Math (both rated green). This study sought to assess the effects of adopting Eureka Math or Into Math on student achievement and to determine whether one of the two curricula was more effective than the other.

Overall effects of switching to a focal curriculum

Overall, switching to Eureka Math or Into Math was primarily effective in raising students' math course grades, with improvements in students' math GPA occurring in every year of implementation. It is unclear whether these improvements reflect increases in students' math knowledge and skills or reflect changes or differences in schools' grading practices. We found little evidence that switching curricula led to improvements in students' scores or proficiency on either the NWEA MAP or the state test, particularly in later years of implementation when grades continued to improve. If the effects on grades reflected growth in student learning, we might have expected students' test performance to also show improvement over time. In a sensitivity analysis, we found that the effects on course grades were detected specifically when comparing schools that switched curricula to schools that continued using the same curricula during the study period—that is, when we compared schools that switched curricula to other schools that would go on to switch in the future, we did not detect the same large, positive effects on math course grades. Together with the absence of impacts on students' test performance, the results of this sensitivity analysis suggest that the effects on course grades may be due to other factors beyond increases in student learning.

Effects over time of switching to a focal curriculum

Results from the Bayesian analysis indicate that the probability of positive effects for both curricula decreased with time after the curricula were adopted. Time may affect the efficacy of curricula in different ways. For example, teachers may need time to learn how to effectively implement and adapt the curricula, making the curricula more effective over the years. In the context of this study, there is an additional reason to think implementation could have improved over time: For over half of schools, their first year of implementation coincided with the COVID-19 pandemic and school closures. However, the results indicate that the probability that the curricula were effective *decreased* over time for all outcomes. This could occur if, in the first year of implementation, teachers were most faithful to the curricula, as they engaged in professional learning and received implementation supports; over time, teachers may adapt the curricula more in ways that are not productive. Few curricular efficacy studies have examined the effects on student outcomes over time. However, two examples—Agodini et al. (2009) and Koedel et al. (2017)—have found that curricular efficacy can increase, decrease, or stay the same over time.

Comparing Eureka Math versus Into Math

In terms of whether one of the two focal curricula was more promising than the other, on average, we found that both had similar effects on student achievement outcomes, though there was some indication that Eureka Math may have been more effective in the first year of implementation. For example, the effect of switching to Eureka Math on NWEA MAP math scores was very likely positive after one year (81 percent probability of an effect greater than zero). However, as noted above, the probability of positive effects decreased with time, approaching even odds after two years (44 percent) and becoming most likely negative after three years (11 percent). Similarly, in the first year of implementation, Eureka Math likely led to improvements in students' proficiency on the state test (86 percent probability of an effect greater than zero), though again, this likelihood fell steadily in each year of implementation. Despite the more promising results for Eureka Math in Year 1, the survey data (which, for almost test all schools, corresponded to their second year of implementation) suggest that otherwise similar students had less positive beliefs about math when taught using Eureka Math than Into Math, and teachers using Eureka Math were less likely to feel the curriculum was appropriate for their students and felt less confident in their ability to meet their students' needs. Interviews with district and school leaders also suggested that Eureka Math may be more difficult for some teachers to implement successfully. In fact, whereas Into Math scored above 90 percent for usability on EdReports, Eureka Math received a score of 66 percent.

Critical factors beyond the focal curricula

Although the research literature shows that adopting a new, quality curriculum *can* improve outcomes for students, the inconsistency of findings across the body of evidence suggests that selecting and implementing an effective curriculum that is right for a given context is a complex process. As noted in guidance by the National Research Council (2004), in the absence of multiple rigorous studies that consistently demonstrate improvements in student learning across different contexts, it is unclear whether adopting a particular curriculum should be seen as an effective intervention to improve student outcomes. It may be that other factors related to math instruction are more significant or consistent levers for improving student learning. For example, in this study we found that teachers' perceptions of the consistency, specificity, and normative authority of their curricula were not correlated with their students' math achievement, whereas their views of the collective efficacy of the teachers in their school and the supportiveness of their school leadership were strong predictors of student outcomes. Although these relationships are not causal, they are consistent with a larger evidence base that points to the importance of effective teachers and school leaders as key contributors to student learning (for example, see Aaronson et al., 2007; McCaffrey et al., 2009; Nye et al., 2004; Rivkin et al., 2005; Rockoff et al., 2004; Blazar & Kraft, 2017; Branch et al., 2012; and Walsh & Dotter, 2020).

Our findings, taken together with other studies of curricular efficacy, suggest the curriculum that is most effective in one context may not be right in another. Curricular efficacy studies examining the same curriculum in different contexts have sometimes found different results. Results can differ even when the comparison curricula are the same. Blazar et al. (2020) found that in California, Eureka Math had a negative impact on student achievement compared to enVision Math, whereas the two curricula were found to be similar when looking at all six states in the study. As another example, whereas Agodini et al. (2009) found that Saxon Math was more effective than the Scott Foresman-Addison Wesley (SFAW) curriculum, Bhatt and Koedel (2012) obtained the opposite result. The authors of the later study note that the earlier study took place in schools with much higher shares of low-income students and students of color and suggest that SFAW may not be the right choice for schools serving these populations. Our study also suggests the efficacy of the curricula may depend on the context. For example, exploratory subgroup results indicate the curricula may have been relatively less effective in raising the performance of Black and Hispanic students on the state standardized test compared to students of other races and ethnicities.

The results of this study should be viewed as being specific to the place and time in which the curricula were implemented. As noted earlier, over half of the schools that switched curricula in the district did so in the 2020–2021 school year—the middle of the COVID-19 pandemic; over a third (the Eureka Math early adopters) switched in 2019–2020; and the remainder switched in 2021–2022. Given the challenges the pandemic presented, it is possible that, under more normal conditions, the effects of adopting the focal curricula could be different. Nevertheless, we do not find evidence that the effects differ depending on the year that schools switched curricula, and as mentioned above, there is also no evidence that the effects improved in later implementation years. Still, we cannot rule out that adopting a new curriculum in the period of 2020–2021 to 2021–2022 presented unique challenges that may have affected the change's ability to positively influence student achievement outcomes, even two to three years later.

Implications for future research

Although additional studies of curricular efficacy can help build the evidence base for a given curriculum, it is worth noting that many potentially high-quality curricula are available and that curricula frequently evolve as publishers develop new versions or new curricula altogether. Currently, EdReports has reviewed 79 math curricula for K–8 students, 38 of which received green ratings. The majority (57 percent) of U.S. school districts adopt a new curriculum every six to 10 years, with an additional 14 percent replacing their curricula every one to five years (Allen & Seaman, 2017). Given the importance of understanding whether a curriculum is effective over time, it may be difficult for research on the efficacy of individual curricula to keep up with the pace of curriculum development and adoption. Therefore, in addition to assessing the efficacy of individual curricula, future research should seek to identify the features of effective curricula, as well as the features of teachers' enactment and the implementation context that make curricula more effective. This is one of the goals of the ongoing AMS project. As noted by Steiner (2017), we know very little about what makes a curriculum effective. Evidence on the features of effective curricula and the implementation factors that support greater efficacy may be more generalizable than individual efficacy studies and may help schools select curricula that are most likely to be effective in their contexts.

Efficacy analysis in 2024

The AMS project collected an additional year of administrative data from the school district in fall 2023, which enabled us to assess how the effects of the curriculum switch evolved after a third and fourth year of implementation. We analyzed an additional year of student and teacher survey data to examine whether the patterns observed in this report remain consistent. The findings from the additional analysis is available in Appendix C of this report.

This page has been left blank for double-sided copying.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135. https://doi.org/10.1086/508733

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., (2009). Achievement effects of four early elementary school math curricula. National Center for Education Evaluation and Regional Assistance, U.S. Department of Education, Institute of Education Sciences. NCEE 2009-4052.

Allen, I. E., & Seaman, J. (2017). What we teach: K-12 school district curriculum adoption process, 2017. Babson Survey Research Group.

Amos, L., Gorang. E, Haderlein, S., Hallgren, K., Hartog J., Johnson, A., Pace, S., Smith, E., Stone, R., & Wood, M. (2022a). Analysis of Middle School Math Systems (AMS) interim report. Princeton, NJ: Mathematica.

Amos, L., Gorang, E., Haderlein, S., Hallgren, K., Johnson, A., Pace, S., Tiwari, N., & Venezia, F. (2022b). Analysis of Middle School Math Systems (AMS): Further investigation into student experiences, instruction, and context. Princeton, NJ: Mathematica.

Angrist, J.D., & Pischke, J.S. (2008). *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.

Baker, A., Larcker, D., & Wang, C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics, 144*(2), 370–395.

Balfanz, R., & Byrnes, V. (2006). Closing the mathematics achievement gap in high-poverty middle schools: Enablers and constraints. *Journal of Education for Students Placed At Risk*, *11*(2), 143–159.

Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis, 34*(4), 391–412.

Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review, 34*, 107–121.

Blazar, D., Heller, B., Kane, T.J., Polikoff, M., Staiger, D.O., Carrell, S., Goldhaber, D., Harris, D.N., Hitch, R., Holden, K.L., & Kurlaender, M. (2020). Curriculum reform in the Common Core era: Evaluating elementary math textbooks across six U.S. states. *Journal of Policy Analysis and Management, 39*(4), 966–1019. https://doi.org/10.1002/pam.22257

Blazar, D., & Kraft, M.A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146–170. https://doi.org/10.3102/0162373716670260

Branch, G.F., Hanushek, E.A., & Rivkin, S.G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals. National Bureau of Economic Research. https://www.nber.org/papers/w17803

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education and Development, 13*(1), 6–33.

Deke, J., Finucane, M., & Thal, D. (2022). The BASIE (BAyeSian Interpretation of Estimates) framework for interpreting findings from impact evaluations: A practical guide for education researchers. (NCEE 2022-005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/2022005/

Desimone, L.M., & Garet, M.S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education, 7*(3), 252–263.

Elmore, R.F., Forman, M.L., Stosich, E.L., & Bocala, C. (2014). The internal coherence assessment protocol & developmental framework: Building the organizational capacity for instructional improvement in schools. Strategic Education Research Partnership.

Gold, T., Carroll, K., Newby, L.D.T., & Steel King, M. (2023). Rounding up: An analysis of math curriculum effectiveness studies. Washington, D.C.: Bellweather. https://files.eric.ed.gov/fulltext/ED629262.pdf

Imbens, G.W., & Wooldridge, J.M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature., 41*(1), 5–86.

Koedel, C., Li, D., Polikoff, M.S., Hardaway, T., & Wrabel, S.L. (2017). Mathematics curriculum effects on student achievement in California. *AERA Open, 3*(1), 1–22.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*(4), 572–606. https://doi.org/10.1162/edfp.2009.4.4.572

Moyer, J.C., Cai, J., Wang, N., & Nie, B. (2011). Impact of curriculum reform: Evidence of change in classroom practice in the United States. *International Journal of Educational Research, 50*(2), 87–99.

National Research Council (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11025

National Center for Education Statistics (NCES). (2023). Education demographic and geographic estimates. http://elibrary.pcu.edu.ph:9000/digi/NA02/2004/11025.pdf

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257. https://doi.org/10.3102/01623737026003237

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458. https://www.jstor.org/stable/3598793

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252. https://doi.org/10.1257/0002828041302244

Schochet, P. (2022). Statistical power for treatment effects using difference-in-differences and comparative interrupted time series estimators with variation in treatment timing. *Journal of Educational and Behavioral Statistics*, 47(4), 367–405. https://doi.org/10.3102/10769986211070625

Slavin, R.E., Lake, C., & Groff, C. (2008). Effective programs in middle and high school mathematics: A best evidence synthesis. Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education.

Steiner, D. (2017). Curriculum research: What we know and where we need to go. Standards Work. https://standardswork.org/wp-content/uploads/2017/03/sw-curriculum-research-report-fnl.pdf

Stone, R., Rosendahl, L., Johnson, A., & Harris, B. (2023). Analyzing middle school mathematics curricula: A comparative study using three measurement tools. Princeton, NJ: Mathematica.

Tarr, J.E., Reys, R.E., Reys, B.J., Chávez, Ó., Shih, J., & Osterlind, S.J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247–280.

Tran, S.T., Majumdar, R., Hu, A.H., & Meyer, J.P. (2022). Default cut scores for MAP Growth reading and mathematics assessments. NWEA Research Report. https://www.nwea.org/uploads/2022/07/MAP-Growth-Default-Cut-Scores_NWEA_linking-study.pdf

Turner, E.E., Drake, C., McDuffie, A.R., Aguirre, J., Bartell, T.G., & Foote, M.Q. (2012). Promoting equity in mathematics teacher preparation: A framework for advancing teacher learning of children’s multiple mathematics knowledge bases. *Journal of Mathematics Teacher Education, 15*(1), 67–82.

Walsh, E., & Dotter, D. (2020). The impact on student achievement of replacing principals in District of Columbia public schools. *Education Finance and Policy, 15*(3), 518–542. https://doi.org/10.1162/edfp_a_00279

Appendix A

Additional Technical Details

This appendix presents additional technical details for the efficacy analyses used to assess the effects of schools switching to Eureka Math or Into Math, and for the descriptive analyses of survey data used to explore differences in the beliefs of students and teachers in schools that used Eureka Math or Into Math.

Technical details of efficacy analyses

Below, we describe the main analytic model used in the efficacy analyses, along with details about the weights used in these analyses, how we treated missing data, and how we constructed the sample. We then describe the analytic models used to conduct subgroup and sensitivity analyses. Finally, we discuss the statistical power of the main analyses and the approach to Bayesian interpretation of the results.

Main analytic model

To estimate the effect of switching curricula on students' math achievement, the study team used the following model:

$$(1)\ \ Y_{ist} = \beta_0 + \textstyle\sum_y \beta_y (P_s \times D_y) + P_s + D_y + \alpha_s + \delta_t + X_i + Z_{it} + \varepsilon_{ist},$$

where Y_{ist} is the outcome for student i in school s and year t; β_0 is an intercept term; P_s is an indicator equal to one if school s ever switched curricula; D_y are indicators equal to one for each year post implementation, $y \in \{1,2,3\}$, such that the calendar year designated as the first implementation year can vary by school depending on the year in which they switched curricula; β_y are the coefficients of interest and represent the difference in trends for schools that switched curricula in implementation year y; α_s are school fixed effects; δ_t are year fixed effects; X_i are characteristics that are stable within students over time, and Z_{it} are time-varying student characteristics; and ε_{ist} is an error term.

To estimate the effects of switching to Eureka Math and Into Math separately, the study team used the following model:

$$(2)\ \ Y_{ist} = \beta_0 + \textstyle\sum_y \theta_y (E_s \times D_y) + \sum_y \rho_y (IM_s \times D_y) + E_s + IM_s + D_y + \alpha_s + \delta_t + X_i + Z_{it} + \varepsilon_{ist},$$

where E_s is an indicator equal to one if school s ever switched to Eureka Math and IM_s is equal to one if school s ever switched to Imagine Math; θ_y and ρ_y are the coefficients of interest and represent the difference in trends for schools in implementation year y for schools that switched to Eureka Math and Into Math, respectively; and all other terms are defined as in Equation 1.

The study team estimated Equations 1 and 2 using ordinary least squares and Huber-White sandwich standard errors to account for clustering at the school level. The models controlled for both time-varying and time-invariant student-level covariates, summarized in Table A.1.

Table A.1. Student-level covariates

Covariate	Description
Time-invariant characteristics	
Race or ethnicity	Mutually exclusive binary indicators of whether student identified with one of the following groups: • Black • Hispanic • American Indian or Alaska Native • Native Hawaiian or Pacific Islander • Two or more races • White
Gender	Binary indicator of whether student identified as female or male
Time-varying characteristics	
English language learner status in year y	Binary indicator of whether student was considered an English language learner in each year
Special education status in year y	Binary indicator of whether student had special education needs in each year
Number of suspensions in year y	Integer count of in-school or out-of-school suspensions student received in each year
Attendance rate in year y	Percent of enrolled days student attended school in each year
State test English language arts score in year y	Student's standardized score on a state-developed summative assessment designed to measure English language arts achievement based on the state's learning standards. We converted scale scores into z-scores by subtracting the average scale score for students in the school district within the same school year and grade level and dividing by the corresponding standard deviation. Scores were imputed for the 2019–20 school year (see section on treatment of missing data below).
NWEA MAP reading score in year y	Student's standardized score on an interim, state-aligned adaptive assessment designed to measure student reading achievement each year and growth over time. We converted scale scores into z-scores by subtracting the average scale score for students in the school district within the same school year and grade level and dividing by the corresponding standard deviation.

Weights

The analysis weighted the student-level data so that each school contributed equally to the results, regardless of the number of students in the school. An individual i in school s was weighted by $W_{is} = {}^{1}/_{N_s}$, where N_s is the number of individuals with nonmissing values of the outcome for school s. These weights ensured the results were not overly influenced by the effects of switching curricula in larger schools. The results therefore reflect the effects of switching curricula for the average school.

Treatment of missing data

The analysis included only students who had nonmissing values of the outcome variables; individuals with missing values of an outcome variable were excluded from the estimation of effects on that outcome. Individuals were also excluded from the analysis if they had missing covariate values. In most cases where a student was excluded for missing covariates, the student had scores for one test but was missing the other test. For example, students with NWEA MAP math outcomes who were missing the state test English languages arts score were excluded, or vice versa.

For state test scores, no data were available in the 2019–20 school year because of the test's cancellation. The estimates of the effects of switching curricula on state test scores do not include observations from students in 2019–20. However, other models included state test English language arts scores as a covariate. The study team imputed these scores in 2019–20 as the average score in non-missing years for each student. If students did not have state test English language arts scores in other years, the value was imputed as the average of other students in the same school in spring 2019, the final year that was a baseline year for all schools.

Analytic sample

The school district provided data from 64 middle schools. The study team excluded one school from the analyses because it exclusively provided remote learning, two schools because they used both focal curricula during the study period (one school switched from Eureka Math to Into Math after two years, and another switched from Into Math to Eureka Math after one year), and five schools because they opened during the baseline period and therefore did not have data available for all baseline years. The study team defined the analytic sample to be students who were enrolled in one of the remaining 56 study schools and who had nonmissing outcome data.

Estimation of effects for subgroups

As an exploratory analysis, we estimated the effect of switching curricula on different subgroups of students based on (1) race or ethnicity or (2) gender. Table A.2 describes how these subgroups were defined. To estimate effects for subgroups of students, we estimated a modified version of Equation 2 that adds an indicator for being in the subgroup and an interaction between that indicator and the curriculum switch indicators. That is, the study team estimated the following model:

$$\text{(3)}\quad Y_{ist} = \beta_0 + \textstyle\sum_y \mu_y\left(g \times E_s \times D_y\right) + \sum_y \pi_y\left(g \times IM_s \times D_y\right)$$

$$+\sum_y \theta_y\left(E_s \times D_y\right) + \sum_y \rho_y\left(IM_s \times D_y\right) + g \times E_s + g \times IM_s + \sum_y\left(g \times D_y\right)$$

$$+E_s + IM_s + D_y + g + \alpha_s + \delta_t + X_i + Z_{it} + \varepsilon_{ist},$$

where g is an indicator for being part of the subgroup of interest. The coefficients θ_y and ρ_y represent the effects of switching to Eureka Math and Into Math, respectively, for students who are *not* part of the subgroup of interest. The coefficients μ_y and π_y represent the differential effect of switching to Eureka Math and Into Math for students who are part of the subgroup of interest relative to students who are not part of the subgroup (the reference group). For example, for male and female students, the effect of switching to Eureka Math after one year for male students would be captured by θ_1, the effect for female students would be represented by the sum $\theta_1 + \mu_1$, and $\mu_1$would represent the difference between female and male students, where a positive value would indicate the effect was greater for girls than boys.

Table A.2. Subgroups examined

Subgroups	Definition
Race and ethnicity	
Black or Hispanic	The student identified as either Black or Hispanic
Not Black or Hispanic	The student identified as a different race or ethnicity other than Black or Hispanic (e.g., White, Asian, Pacific Islander, etc.)
Gender	
Female	The student identified as female
Male	The student identified as male

Sensitivity analyses

Additional analyses explored whether estimates were sensitive to the timing of when schools adopted their new curriculum. To understand whether the timing of switching to a new curriculum mattered, the study team used a modified version of Equation 1 that added indicators, P_{st}, for whether a school switched to a focal curriculum in the school year where the spring semester was in $t \in \{2021, 2022\}$, as well as the interaction of these indicators with the switch year:

$$(4)\quad Y_{ist} = \beta_0 + \sum_y \beta_y \left(P_s \times D_y\right) + \sum_{t=2021}^{2022} \sum_y \beta_{yt} \left(P_{st} \times D_y\right)$$

$$+P_s + \sum_{t=2021}^{2022} P_{st} + D_y + \alpha_s + \delta_t + X_i + Z_{it} + \varepsilon_{ist}.$$

The coefficients β_y represent the effects of switching to a focal curriculum for schools that switched in the 2019–20 school year. The coefficients β_{yt} represent the differential effect of switching to a focal curriculum for schools that switched in the 2020–21 or 2021–22 school years relative to schools that switched in 2019–20. For example, the effects of switching to a focal curriculum after one year for students in schools that switched in 2019–20 is captured by β_1, the effect for students in schools that switched in 2020–21 is represented by the sum $\beta_1 + \beta_{1,2021}$, and $\beta_{1,2021}$ represents the difference between the two groups.

To continue examining the importance of timing, the study team also examined whether the estimates were sensitive to excluding schools that were early adopters of the focal curricula. To do this, the study team estimated the model in Equation 2 with two samples: (1) schools that switched in the 2020-21 school year and schools that never switched to a focal curriculum and (2) all schools except for those that switched in the 2019–20 school year.

In addition, further analyses examined whether estimates of the effects of switching to a focal curriculum were sensitive to the composition of schools that made up the comparison group. The comparison group includes a mix of counterfactual conditions. To isolate the different counterfactuals, the study team estimated a modified version of Equation 1 that only looks at the effects of switching to a focal curriculum one year after the switch:

$$(5)\quad Y_{ist} = \beta_0 + \beta_1 (P_s \times D_1) + P_s + D_1 + \alpha_s + \delta_t + X_i + Z_{it} + \varepsilon_{ist},$$

where P_s is an indicator for a treatment school that switched curriculum and D_1 is an indicator for being one year post implementation, similar to Equation 1. The samples for these analyses only included treatment schools that switched to the same curriculum in the same year. The analysis then isolated the comparison schools to one of two groups: (1) schools that never switched curricula or (2) schools that switched to the same curriculum as the treatment schools but did so after the treatment schools. Table A.3 lists the sample restrictions for each of these sensitivity analyses, including which schools were in the treatment and comparison groups, and which years of data were included.

Table A.3. Sample restrictions for sensitivity analyses to isolate counterfactual conditions

Treatment schools	Comparison schools	Years of data
Treatment schools that switched to Eureka Math		
Switched in 2019–20	Schools that never switched curricula	2015-16 through 2019–20
Switched in 2019–20	Schools that switched to Eureka Math in 2020–21 or 2021–22	2015-16 through 2019–20
Switched in 2020–21	Schools that never switched curricula	2015-16 through 2020–21
Switched in 2020–21	Schools that switched to Eureka Math in 2021–22	2015-16 through 2020–21
Switched in 2021–22	Schools that never switched curricula	2015-16 through 2021–22
Treatment schools that switched to Into Math		
Switched in 2020–21	Schools that never switched	2015-16 through 2020–21
Switched in 2020–21	Schools that switched to Into Math in 2021–22	2015-16 through 2020–21
Switched in 2021–22	Schools that never switched curricula	2015-16 through 2021–22

Statistical power

Table A.4 describes the parameters used to examine the statistical power of the main analysis model in Equation 2. The minimum detectable effect sizes (MDES) are in standard deviation units, and minimum detectable effects (MDE) are in the same units as the outcome. The table indicates that analyses that examine math course grades have adequate statistical power, with a MDE that is roughly equal to the effect sizes estimated in Table IV.2. The analysis for math course passing rates is moderately powered, defined as having a MDE between 100 and 150 percent of the expected effect. The other analyses are in danger of being underpowered, which the study team defines as a MDE that is more than 150 percent of the expected effect. For example, the MDE for NWEA MAP math z-scores is 180 percent of the largest estimated effect reported in Table IV.2 for that outcome.[25]

The failure to reject the null hypothesis of no effect for these outcomes can therefore reflect either effects that are not statistically different from zero or insufficient power to detect differences smaller than the MDEs. Although this limits the main analyses, we supplement the traditional frequentist analyses with Bayesian analyses, described in the next section.

---

[25] These power calculations are based on a standard difference-in-difference framework that does not take into account the staggered timing of when schools switched curricula. Recent work suggests that accounting for staggered timing as well as the autocorrelation of errors within schools across time likely reduces the minimum detectable effects and effect sizes compared to what we report in Table A.4 (Schochet, 2022).

Table A.4. Parameters and minimum detectable effect sizes from power calculations

	NWEA MAP math		State test math		Math course grades	
	Proficient (%)	z-score (mean)	Proficient (%)	Proficient (%)	z-score (mean)	Proficient (%)
Parameters						
Number of individuals	37,249	37,249	31,124	31,124	36,993	36,993
Standard deviation of the outcome	1.00	0.22	1.00	0.37	1.15	0.50
Intraclass correlation coefficient	0.11	0.06	0.13	0.10	0.09	0.06
Proportion of within-group variance	0.05	0.00	0.04	0.02	0.09	0.03
Proportion of group-level variance	0.65	0.20	0.59	0.38	0.41	0.29
Proportion of variance of treatment explained by covariates	0.07	0.07	0.07	0.07	0.07	0.07
Proportion of outcome variance explained by persistent factors	0.84	0.61	0.67	0.58	0.62	0.46
Minimum detectable effect size and effect						
MDES	0.18	0.30	0.29	0.36	0.32	0.34
MDE	0.18	0.07	0.29	0.13	0.36	0.17

Bayesian interpretation

We use the BASIE tool developed by Deke et al. (2022) to combine the frequentist estimates from the main analysis with a prior distribution to estimate a Bayesian posterior distribution. The prior distribution is based on an analysis of all estimates in the What Works Clearinghouse (WWC) database that meet standards, with or without reservations. To align the prior distribution more closely to the intervention being studied, the study team limited the prior distribution to interventions in middle grades that examined math achievement outcomes. In addition, the prior distribution includes adjustments for small studies, which are susceptible to more statistical noise, and is not centered at zero to reflect the actual findings in the WWC, which tend to be small but positive, on average.

It is important to note that statements made using the BASIE tool can only be interpreted relative to the selected prior distribution and are not predictive statements about the effects in the future, but instead retrospective statements about the effect of switching curricula in the context of this study.

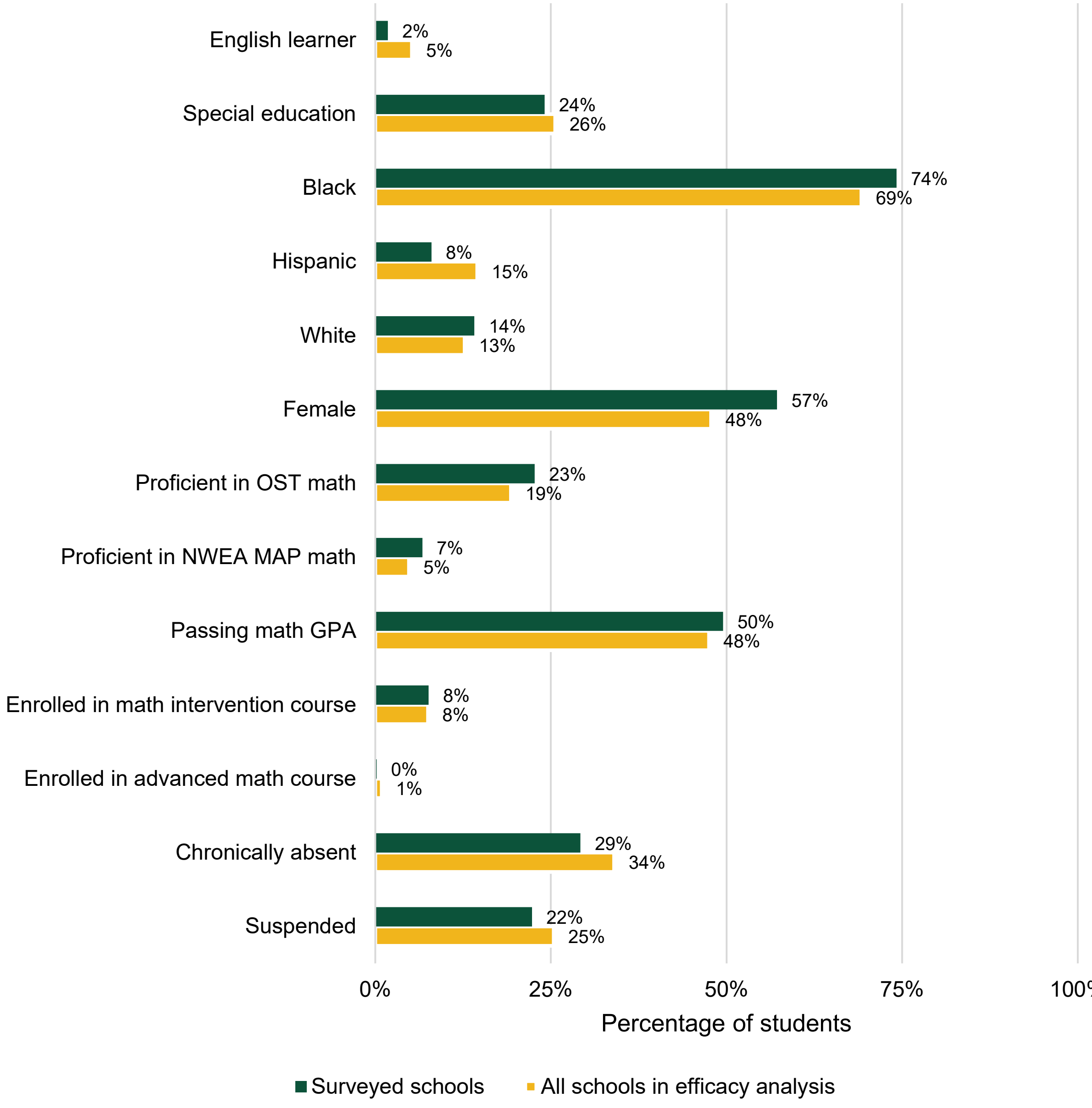
Technical details of survey analyses

Below, we describe details about the student and teacher survey samples and the survey constructs we analyzed.

Survey samples

To create the sample of schools for study participation and data collection activities, the study team worked with each of the four participating districts to identify and recruit a set of schools with grades 6 to 8 that were using the curricula of interest and were willing to participate. This resulted in 10 participating schools in the school district. Of these 10 schools, six of them adopted Eureka Math (five schools switched to Eureka Math in 2020–21 and one switched in 2021–22) and four of them adopted Into Math (all switched to Into Math in 2020–21). Figure A.1 compares the average characteristics of students in the 10 surveyed schools versus all 56 schools in the efficacy analysis.

Figure A.1. Characteristics of students in surveyed schools compared to all schools in the efficacy analysis



Source: School district data.

Note: None of the differences between schools is statistically significant at the 5 percent level.

Within these schools, to identify the classrooms, teachers, and students for participation in our surveys, we obtained a list of all middle school math classrooms from each study school. For each classroom, we received information such as grade (6, 7, or 8), teacher name, class level (for example, below grade level, general education, advanced), and number of students. We excluded classrooms designated as below grade level or advanced, as well as classrooms with fewer than 12 students. Among the remaining classrooms, we asked all math teachers to complete a survey. We randomly selected one classroom per grade per school, for a total of three classrooms per school, from which we collected student surveys.

To understand teachers' and students' experiences with math classrooms, we administered surveys to both teachers and students in the fall and spring of the 2021–22 school year. For nine of the 10 schools in the sample, this school year corresponded to their second year of implementation of the new curricula.

We administered the teacher survey to 21 teachers in the school district. The fall survey had a 67 percent response rate, and the spring survey had a 57 percent response rate (Table A.5). The fall teacher survey collected information on teachers' teaching background and experience, perceptions and use of the math curriculum, and teaching practices. The spring teacher survey collected information about use of the math curriculum and teachers' experiences with professional learning. Each survey was administered via a web-based platform and took approximately 30 minutes to complete.

We administered the student survey to 731 students in the school district. The fall survey had a 69 percent response rate, and the spring survey had a 42 percent response rate (Table A.5). The student survey included questions related to student beliefs (growth mindset, achievement identity, math persistence, math self-efficacy, and math enjoyment), student engagement, and student interest and persistence in math. Both fall and spring administrations used the same instrument. The student survey took approximately 10 minutes to complete. Schools were offered the option of administering the survey by web or a paper version, with the majority of schools selecting the latter.

Table A.5. Teacher and student survey samples, completions, and response rates

Survey	Eligible sample	Completions	Response rate
Fall student survey	731	501	69%
Spring student survey	731	306	42%
Fall teacher survey	21	14	67%
Spring teacher survey	21	12	57%

Source: Survey data.

Survey constructs

The study team constructed scales to calculate a respondent's aggregate score across items capturing a single construct. We constructed scales by taking the mean across all items on that scale (after reverse coding any negatively worded items to ensure higher scores consistently indicated more positive beliefs). We calculated Cronbach's alpha to assess the reliability of each scale. First, we ensured that the alpha for each scale was equal to 0.70 or greater and that the alpha value would not be improved by removing any items. If either of these conditions was not met, the study team discussed as a group whether to remove any items from the scale. A list of all the scales created from the student and teacher surveys, their range of values, definitions, and the survey items included in those scales are listed in Tables A.6 and A.7.

Table A.6. Definitions and items included in the student survey scales

Construct	Range of values	Construct definition	Survey items included[26]
Student engagement	1 to 5	Positive and active participation in math class including the desire to meet academic expectations (such as earning good grades and test scores), comply with social and behavioral classroom norms (such as being a good small group partner), engage cognitively (such as the personal drive or commitment to improve conceptual understanding of a particular math topic), and engage emotionally (such as being excited when playing math games)	When reading the following statements, think about your current math class and decide how well the statements describe you. a. I don't think that hard when I am doing work for math class. b. I complete my math homework on time. c. I don't participate in math class. d. I do other things when I am supposed to be paying attention. e. I try to work with others who can help me in math. f. I build on others' ideas. g. I try to understand other people's ideas in math class. h. I don't care about other people's ideas. When reading the following statements, think about your current math class and decide how well the statements describe you. a. I try to understand my mistakes when I get something wrong. b. I want to understand what is learned in math class. c. I try to help others who are struggling in math. d. I talk about math outside of class. e. I think that math class is boring. f. I don't like working with classmates.
Math enjoyment	1 to 5	The belief that doing math and being in math class is fun	When reading the following statements, think about your current math class and decide how well the statements describe you. a. I look forward to math class. b. I enjoy learning new things about math. c. I feel good when I am in math class. d. I often feel frustrated in math class. e. I don't care about learning math. f. I don't want to be in math class. g. I often feel down when I am in math class. h. I get worried when I learn new things about math.

---

[26] The letters in the "Survey items included" column represent the actual survey item letters from the student surveys.

Math self-efficacy	1 to 6	Students’ confidence in solving math problems and performing math-related tasks	When reading the following statements, think about your current math class and decide how well the statements describe you. How much do you disagree or agree with the statements below? b. I learn things quickly in math. c. I am good at working out difficult math problems. f. I believe that I can be successful in my math class. h. I am confident that I can understand the material in my math class. i. I know I can learn the materials in my math class.
Achievement identity	1 to 6	Students identifying and holding a self-concept as someone who can achieve academically	When reading the following statements, think about your current math class and decide how well the statements describe you. How much do you disagree or agree with the statements below? a. I usually do well in math. b. Math is harder for me than any other subject. c. My teacher tells me I am good at math. How much do you disagree or agree with the statements below? a. My classmates think I am good at math. b. My friends think I am good at math. c. My parents think I am good at math.
Growth mindset	1 to 6	Students’ belief that their ability to learn is not fixed but can be developed over time	When reading the following statements, think about your current math class and decide how well the statements describe you. How much do you disagree or agree with the statements below? a. Being a top math student requires a special talent that just can’t be taught. b. If you want to succeed in math, hard work alone just won’t cut it; you need to have a natural gift or talent. c. When you have to try really hard in math in school, it means you can’t be good at math. d. Being a ”math person” or not is something that you really can’t change. Some people are good at math and other people aren’t.

Table A.7. Definitions and items included in the teacher survey scales

Construct	Range of values	Construct definition	Survey items included[27]
Perceptions of curriculum and professional learning			
Curriculum consistency	1 to 5	Teacher perceptions of the coherence or alignment of a curriculum with content standards, assessments, and instructional policy (curriculum consistency)	Please mark the extent to which you disagree or agree with each of the following statements. c. Promotes consistency of instruction among math classes at the same grade-level. d. Promotes continuity of math instruction between grades. e. Is aligned to my state's math content standards. f. Is aligned to my state's math assessments. g. Is aligned to my district or school's math formative or summative assessments.
Curriculum specificity	1 to 5	Teacher perceptions of the level and clarity of detail a curriculum developer provides on instructional content, resources, assessment tasks, and learning objectives (curriculum specificity) influence adherence to an intended curriculum	Please mark the extent to which you disagree or agree with each of the following statements. a. Clearly indicates the math content teachers should teach. b. Provides detailed guidance on math objectives. c. Provides the sequence in which topics are covered in math. d. Sets the pace for covering topics in math. e. Provides criteria for grading students in math. f. Recommends evaluation and assessment activities in math. g. Recommends books and other materials that support math instruction. h. Is easy to use and implement. i. Helps me plan instruction.
Curriculum normative authority	1 to 5	The extent to which teachers feel the curriculum is appropriate for their students	Please mark the extent to which you disagree or agree with each of the following statements. [Curriculum]: k. Is culturally relevant. l. Is engaging for students. Please mark the extent to which you disagree or agree with each of the following statements. [Curriculum]: a. Is too inflexible for teachers to effectively teach students. b. Includes more content than can be covered adequately in the school year.

---

[27] The letters in "Survey items included" represent the actual survey item letters from the teacher surveys.

Construct	Range of values	Construct definition	Survey items included[27]
			c. Is too rigorous for most students I work with. d. Helps prepare students for state standardized tests. e. Excludes important content that students should learn. f. Appropriately addresses the needs of students who are designated as dual language learners (DLLs). g. Appropriately addresses the needs of students who have Individualized Education Programs (IEPs) or the equivalent.
Professional learning coherence	1 to 7	The extent to which teachers feel that professional learning content, goals, and activities are consistent with the school curriculum and goals, teacher knowledge and beliefs, the needs of students, and school, district, and state and policies	Please indicate the extent to which you disagree or agree with each of the following statements. I am engaged in professional learning activities for [Curriculum] that are… a. Integrated/linked with my daily lessons/math curriculum. b. Isolated and/or unconnected to other professional learning activities that I participate in. c. Aligned with my school's mission and goals. d. Consistent with district policies (such as state standardized testing and content standards). e. Aligned with feedback from observations of my teaching.
Professional learning usefulness	0 to 5	The extent to which teachers feel that the professional learning they received helped them improve their math knowledge, instructional skills, and mindsets	Think about your experience with professional learning since the start of the current school year (2021–2022), including the summer of 2021. Please include both in-person and online activities. Please indicate the extent to which your professional learning activities are assisting you in each of the following ways. a. Improving my own content knowledge of math. b. Improving my understanding of how students learn math. c. Advancing my understanding of how to use [Curriculum] in my classroom. d. Providing me with teaching strategies that have improved my math instruction. e. Helping me analyze student performance data in math to identify student needs. f. Having a positive impact on my instruction in math.

Construct	Range of values	Construct definition	Survey items included[27]
			g. Improving my use of math teaching strategies that show respect for the cultural backgrounds of my students. h. Challenging mindsets, expectations, and biases about students to emphasize high expectations for all students. i. Supporting me in being responsive to students' backgrounds, cultures, and points of view. j. Encouraging me to take action when math instructional materials are lacking in representation of multiple perspectives and identities. k. Helping me address the social-emotional needs of my students. l. Helping me make math relevant for my students.
Beliefs about school environment			
Collective efficacy	1 to 7	The extent to which teachers believe that the teachers in their school can collectively improve their teaching to positively influence student learning	Please mark the extent to which you disagree or agree with the following statements. a. Teachers and administrators in our school are interdependent and value each other. b. Staff in our school work together to predict instructional needs and prevent academic issues rather than react and repair. c. Administrators in our school seek to understand instructional problems/issues rather than blame others. d. Teachers in our school seek to understand instructional problems/issues rather than blame others. e. Teachers are empowered to make instructional decisions rather than wait for administrators to tell us what to do. f. Teachers in our school support each other with instructional problems/issues. g. Teachers in our school support each other with classroom management problems/issues. h. Teachers use common planning time to work in units/teams rather than plan instruction as separate individuals. i. Administrators in this school truly believe every child can learn. j. Teachers in this school truly believe every child can learn.

Construct	Range of values	Construct definition	Survey items included[27]
Collective participation	0 to 3	The extent to which teachers believe their professional learning involves opportunities to build an interactive learning community with other teachers (in the same grade, content area or school)	During your collaborative planning time, to what extent do you focus on the following activities: During collaborative planning, we spend time: a. Discussing student work or test scores b. Co-designing lessons c. Reflecting on lessons we taught to identify what went well and what to improve d. Discussing strategies for managing student behavior e. Discussing culturally responsive teaching strategies f. Sharing lesson plans and other instructional resources g. Discussing how to use [Curriculum] h. Discussing learning standards i. Working on individual work, like grading student work or lesson planning j. Discussing strategies for differentiating instruction for special populations k. Other (please specify)
Supportive leadership	1 to 7	The extent to which teachers feel encouraged by school leadership to implement learned knowledge and competencies in the classroom	Next, we would like to know more about your school's leadership. How much do you disagree or agree with each of the following statements? The leadership at this school… a. Makes clear to the staff expectations for meeting instructional goals. b. Communicates a clear vision for our school. c. Sets high standards for teaching. d. Understands how students learn. e. Sets high standards for student learning. f. Carefully tracks student academic progress. g. Knows what's going on in my classroom. h. Actively monitors the quality of teaching in this school. i. Actively encourages or supports me in implementing what I learned into my classroom practice.
Beliefs about own instruction			
Confidence with culturally responsive mathematics teaching	0 to 10	The extent to which teachers feel confident incorporating culturally responsive math teaching practices into their teaching	Please rate your confidence in the following areas on a scale from 0 (not confident at all) to 10 (extremely confident). a. Adapting instruction to meet the needs of my students b. Using a variety of teaching methods c. Using developmentally appropriate practices

Construct	Range of values	Construct definition	Survey items included[27]
			d. Creating positive relationships in the classroom e. Using my students’ cultural backgrounds to help make learning meaningful f. Adapting instructional materials to adequately and appropriately represent cultural groups g. Helping my students feel like important members of the classroom h. Explaining new concepts using examples that are taken from my students’ everyday lives i. Integrating social or political issues into class discussions or assignments j. Uncovering my own implicit biases in my teaching practice k. Supporting my students to be active in social or political causes
Belief in culturally responsive pedagogy	1 to 5	The extent to which teacher beliefs align with teaching practices that recognize that cultural and linguistic differences should be treated as assets for teaching and learning	Please mark the extent to which you disagree or agree with each of the following statements. NOTE: By culturally responsive practice, we mean teaching practices that recognize that cultural and linguistic differences should be treated as assets for teaching and learning. All students, their families, and their communities have cultural capital or knowledge, abilities, and networks that can and should be leveraged in classrooms. a. Culturally responsive practice undermines classroom unity by emphasizing cultural differences. b. Culturally responsive practice is essential for creating an inclusive classroom environment. c. Regardless of cultural differences, all children learn from the same teaching method. d. A color-blind approach to teaching is effective for ensuring respect for all culturally diverse students. e. Encouraging respect for cultural diversity is essential for creating an inclusive classroom environment.

Construct	Range of values	Construct definition	Survey items included[27]
Confidence meeting student needs (self-efficacy)	0 to 10	The extent to which teachers feel confident teaching students at different achievement levels, of diverse backgrounds, and with individualized learning or language learning needs	Please rate your confidence in the following on a scale from 0 (not confident at all) to 10 (extremely confident). a. Teaching math to students performing on grade level b. Teaching math to students performing below grade level c. Teaching math to students performing above grade level d. Teaching math to students who are experiencing poverty e. Teaching math to students who have IEPs or the equivalent f. Teaching math to students who are designated as English learners g. Teaching math to address unfinished learning related to COVID-19
Self-reported use of instructional practices			
Ambitious instruction	0 to 3	The frequency with which teachers report employing ambitious instructional strategies requiring students to explore multiple representations of mathematics, evaluate representations and approaches to mathematical concepts or procedures, and engage critical thinking skills	In how many lessons in a typical math class during a typical week do you have students engage in the following activities? a. Look for and make use of structure (e.g., patterns in numbers, shapes, or algorithms) b. Explain their reasoning or thinking in solving a problem orally or in writing c. Revisit previous grades' content (e.g., content not explicitly included in my grade-level standards) to fill in learning gaps d. Make sense of problems that do not include obvious procedures and persevere in solving them e. Use mathematical language and symbols appropriately when communicating about mathematics. f. Apply mathematics to solve problems in real-world contexts g. Engage in grade-level mathematics for the majority of the classroom time h. Focus on conceptual understanding of the mathematics they are learning

Construct	Range of values	Construct definition	Survey items included[27]
Procedural instruction	0 to 3	The frequency with which teachers report employing teaching practices that are not conceptually sophisticated, with rote student interactions that do not require them to engage critical thinking skills	In how many lessons in a typical math class during a typical week do you have students engage in the following activities? a. Learn or practice basic facts, concepts, and procedures related to a topic e. Answer multiple-choice, fill-in-the-blank, or true/false questions (e.g., worksheets, quizzes, tests, or warmups) h. Receive direct instruction k. Engage in call and response m. Take notes from lectures or the textbook
Culturally relevant mathematics teaching	0 to 3	The frequency with which teachers report incorporating culturally responsive math teaching practices into their teaching	In a typical week for math about how often do you engage in each of the following activities? a. Create opportunities for students to pose authentic questions and/or investigate real-world problems using math b. Create opportunities for students to discuss and explore multiple representations of mathematical concepts and problem-solution paths c. Create opportunities for students to discuss mathematics in meaningful and rigorous ways (e.g., debate mathematics ideas/solution strategies, use mathematics terminology, develop explanations, communicate reasoning, make generalizations) d. Support and scaffold the oral and written academic language development of multilingual students (e.g., gesturing, use of objects [realia], use of cognates, revoicing, graphic organizers and manipulatives) e. Create opportunities for students to draw on their lived experience, local community context, and/or cultural and linguistic heritage as resources for individual and collective learning f. Create opportunities for students to draw connections between math and other content areas g. Create opportunities for students to pose questions about societal challenges of relevance to them and/or instructional tasks that explore, critique, and test solutions to those issues

Appendix B

Additional Results

This page has been left blank for double-sided copying.

This appendix presents the results of additional exploratory and sensitivity analyses.

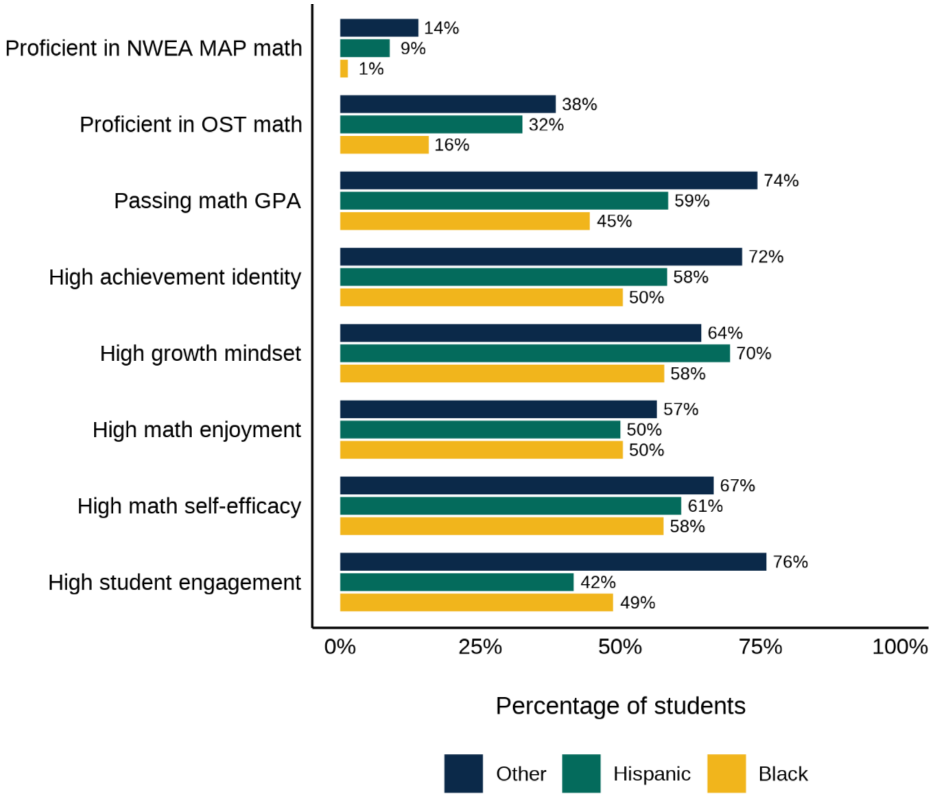
Subgroup analyses

Chapter IV presented the results of subgroup analyses that explored whether the curricula had differential effects for students who (1) identified as either Black or Hispanic compared to students of other races and ethnicities and (2) identified as female compared to students who identified as male. In this section, we describe how average outcomes differed for these subgroups and present additional results.

Descriptive differences in outcomes by subgroup

To put the results of the exploratory subgroup analyses into context, we examined how math achievement and beliefs differed for the different groups of students that were part of these analyses (Figures B.1 and B.2).[28] Compared to students of other races and ethnicities, Black and Hispanic students were less likely to be proficient in math and have a passing math GPA. They were also less likely to report high levels of math achievement identity, enjoyment, self-efficacy, and engagement. While female students were more likely than male students to be proficient in math and have a passing math GPA, they were less likely to report high levels of math enjoyment, math self-efficacy, and math engagement.

Figure B.1. Average math achievement and beliefs, by student race and ethnicity



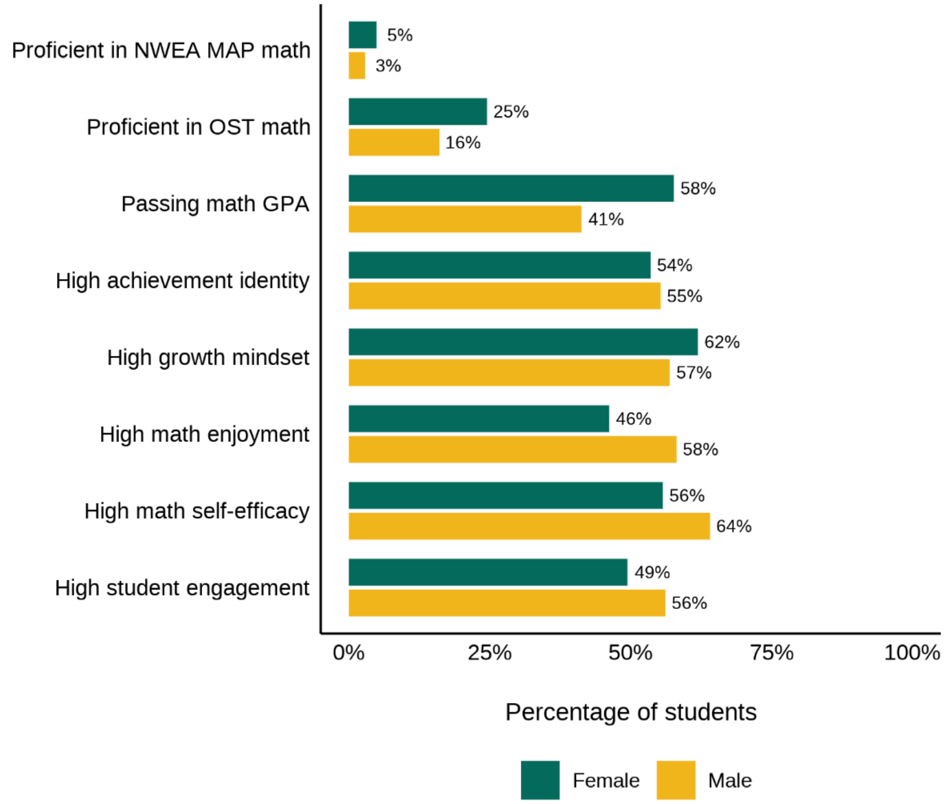
Source: School district and student survey data.

Notes: Only the 10 surveyed schools are included. Achievement data are from 2020–2021 and belief data are from 2021–2022.

---

[28] Because data on beliefs were only available for the 10 schools that participated in survey data collection, all results in Figures B.1 and B.2 are constrained to students in these schools to allow for comparisons across outcomes.

Figure B.2. Average math achievement and beliefs, by student gender



Source: School district and student survey data.

Notes: Only the 10 surveyed schools are included. Achievement data are from 2020–2021 and belief data are from 2021–2022.

Additional results for Black and Hispanic students

Given the relatively smaller population of Hispanic students in the school district, as well as the similarity of the subgroup analyses for Black and Hispanic students, we presented findings for Black and Hispanic students as one group in the main report. In the school district, 64 percent of students identified as Black in the 2018–2019 school year, 17 percent identified as Hispanic, 15 identified as White, and 4 percent identified as other races or ethnicities. Table B.1 reports the results of this same analysis but treats Black and Hispanic students as separate groups. As in Table IV.4, the focal curricula were relatively less effective in raising the performance of Black and Hispanic students on the state standardized test compared to other students, although for Hispanic students, these differences were primarily for Eureka Math and not Into Math.

Table B.1. Effects of switching to Eureka Math and Into Math, by students' race and ethnicity for Black and Hispanic students separately

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Difference in effects for Black students compared to other races or ethnicities						
Eureka Math after 1 year	0.09	-0.03‡	-0.27†	-0.19*	0.17	0.05
	(0.07)	(0.08)	(0.23)	(0.08)	(0.18)	(0.05)
Into Math after 1 year	0.02	-0.01	-0.49*†	-0.16*	0.18	0.07
	(0.08)	(0.08)	(0.23)	(0.08)	(0.19)	(0.07)
Eureka Math after 2 years	0.02	-0.03	-0.22	-0.13	0.24	0.13**
	(0.07)	(0.07)	(0.23)	(0.08)	(0.18)	(0.04)
Into Math after 2 years	-0.04	0.03	-0.27	-0.16*	0.11	0.04
	(0.08)	(0.09)	(0.23)	(0.07)	(0.19)	(0.05)
Eureka Math after 3 years	0.06	-0.02	-0.17	-0.20*	0.12	0.08
	(0.08)	(0.08)	(0.23)	(0.08)	(0.18)	(0.05)
Difference in effects for Hispanic students compared to other races or ethnicities						
Eureka Math after 1 year	0.02	-0.07‡	-0.01	-0.12	0.15	0.08
	(0.12)	(0.09)	(0.12)	(0.07)	(0.15)	(0.09)
Into Math after 1 year	0.06	-0.04	-0.16	-0.08	0.14	0.09
	(0.12)	(0.09)	(0.11)	(0.06)	(0.20)	(0.11)
Eureka Math after 2 years	0.04	-0.06	-0.06	-0.13*	0.30†	0.18†
	(0.12)	(0.09)	(0.11)	(0.06)	(0.16)	(0.09)
Into Math after 2 years	-0.04	-0.02	0.02	-0.11	-0.00†	0.04†
	(0.12)	(0.09)	(0.11)	(0.06)	(0.16)	(0.10)
Eureka Math after 3 years	0.03	-0.06	-0.01	-0.16*	0.20	0.15
	(0.13)	(0.09)	(0.11)	(0.07)	(0.16)	(0.10)
Effects for students of other races or ethnicities (reference group)						
Eureka Math after 1 year	-0.00	0.05	0.18†	0.19*	0.23	0.05
	(0.07)	(0.08)	(0.22)	(0.09)	(0.13)	(0.06)
Into Math after 1 year	-0.02	0.03	0.37†	0.14	0.14	0.02
	(0.08)	(0.08)	(0.21)	(0.09)	(0.16)	(0.08)
Eureka Math after 2 years	-0.01	0.05	0.11	0.09	0.12	-0.02
	(0.07)	(0.08)	(0.21)	(0.08)	(0.15)	(0.06)
Into Math after 2 years	-0.03	-0.01	0.07	0.11	0.22	0.04
	(0.08)	(0.09)	(0.22)	(0.07)	(0.16)	(0.07)
Eureka Math after 3 years	-0.13	0.03	-0.07	0.09	0.19	0.01
	(0.08)	(0.08)	(0.22)	(0.09)	(0.19)	(0.08)
N students	19,704	19,704	18,775	18,775	19,642	19,642
N schools	56	56	56	56	56	56
N years	7	7	6	6	7	7

Source: School district data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses. The top two panels show the estimated differences in the effects of the curricula between (1) the subgroup of interest and (2) the corresponding reference group; a positive (negative) coefficient there indicates the effect was larger (smaller) for the subgroup of interest compared to the reference group. The bottom panel shows the estimated effects for the reference group.

† Difference between Eureka Math and Into Math is statistically significant at the 0.05 level, two-tailed test.

‡ Difference between Black and Hispanic is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Sensitivity analyses

Additional analyses focused on assessing whether the main efficacy findings presented in Table IV.1 and IV.2 were sensitive to (1) the timing of when schools adopted their new curriculum or (2) the composition of schools that made up the comparison group.

Timing of curriculum adoption

To assess whether the timing of when schools switched curricula matters, we tested whether the effects of adopting a focal curriculum differ based on schools' switch year (Table B.2). This analysis does not distinguish whether schools switched to Eureka Math or Into Math, as its focus is on assessing whether the switch year matters. The findings align closely to the main results, with no significant effects of switching on NWEA MAP or state test scores or proficiency rates, large positive effects on math grades and passing rates, and no detectable differences in these estimated effects by schools' switch year.

Table B.2. Effects of switching to a focal curriculum, by switch year

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Effects for schools that switched in 2019–2020						
After one year	0.02 (0.03)	0.00 (0.01)	NA	NA	0.31* (0.13)	0.11* (0.05)
After two years	0.02 (0.04)	0.00 (0.01)	0.02 (0.18)	-0.02 (0.09)	0.32** (0.11)	0.09* (0.05)
After three years	-0.11 (0.08)	-0.00 (0.01)	-0.14 (0.16)	-0.06 (0.04)	0.35** (0.09)	0.13** (0.05)
Difference in effects for schools that switched in 2020–2021						
After one year	0.01 (0.04)	0.00 (0.01)	0.08 (0.16)	0.03 (0.08)	0.13 (0.17)	0.02 (0.06)
After two years	-0.09 (0.06)	0.01 (0.01)	-0.03 (0.11)	0.01 (0.10)	0.10 (0.13)	0.05 (0.05)
Difference in effects for schools that switched in 2021–2022						
After one year	-0.01 (0.07)	0.01 (0.02)	0.03 (0.16)	0.05 (0.03)	-0.05 (0.26)	-0.01 (0.10)
N students	19,704	19,704	18,775	18,775	19,642	19,642
N schools	56	56	56	56	56	56
N years	7	7	6	6	7	7

Source: School district data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses. The first panel shows the estimated effects for schools that switched in 2019–2020 (the reference group). The second two panels show the estimated differences in the effects between schools that switched in later years and the early adopter schools; a positive (negative) coefficient indicates the effect was larger (smaller) for schools that switched later compared to the early adopter group.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

NA = Not applicable. Because the state test was cancelled in 2019–2020, it was not possible to estimate the effects of switching to Eureka Math after one year for schools that switched in 2019–2020.

To assess whether the timing of when schools switched curricula matters, we also estimated the effects of adopting Eureka Math or Into Math excluding early adopter schools. As shown in Chapter III, 19 schools that were early adopters of Eureka Math in 2019–2020 tended to have students with higher average test scores at baseline, among other differences in their student demographics. Although differences in the average characteristics of schools that switched and did not switch to Eureka Math in 2019–2020 should not bias the estimates if the trends in these characteristics were not diverging before the switch, we explore whether the results change when excluding from the analyses the schools that were early adopters.

When estimating the effects of adopting Eureka Math and Into Math including only the 28 schools that switched curricula in 2020–2021 (when most schools made the switch) and the four schools that never switched, we obtain results similar to those in the main analyses (panel 1 of Table B.3). Unlike the main analyses that include all schools, this subsample of schools has the advantage of providing a stable set of treatment and comparison schools in every year of the analysis; in addition, these schools more closely resemble the average district school. When we also include the five schools that switched curricula in 2021–2022, the findings are again similar (panel 2 of Table B.3). Therefore, we find no evidence that the main findings are sensitive to the timing of the curriculum adoption or the potential for selection bias from the unique characteristics of early adopter schools.

Table B.3. Effects of switching to Eureka Math and Into Math, excluding early adopter schools

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Schools that switched in 2020–2021 and that never switched						
Eureka Math after one year	0.09†	0.01	0.10	0.05	0.48**	0.14*
	(0.05)	(0.01)	(0.16)	(0.08)	(0.14)	(0.05)
Into Math after one year	0.00†	0.01	0.06	0.02	0.43**	0.14**
	(0.05)	(0.01)	(0.17)	(0.08)	(0.14)	(0.05)
Eureka Math after two years	-0.03	0.02	0.04	-0.01	0.41**	0.15**
	(0.08)	(0.01)	(0.15)	(0.03)	(0.09)	(0.05)
Into Math after two years	-0.10	0.00	-0.05	-0.01	0.32**	0.11*
	(0.08)	(0.01)	(0.15)	(0.03)	(0.08)	(0.05)
N students	11,845	11,845	11,217	11,217	11,835	11,835
N schools	32	32	32	32	32	32
N years	7	7	6	6	7	7
Schools that switched in 2020–2021 or 2021–2022 and that never switched						
Eureka Math after one year	0.07†	0.01	0.06	0.05	0.41**	0.13*
	(0.06)	(0.01)	(0.16)	(0.05)	(0.13)	(0.06)
Into Math after one year	-0.02†	0.01	0.06	0.03	0.30*	0.11
	(0.06)	(0.01)	(0.16)	(0.05)	(0.13)	(0.05)
Eureka Math after two years	-0.02	0.02	0.05	-0.01	0.49**	0.17**
	(0.07)	(0.01)	(0.15)	(0.03)	(0.12)	(0.06)

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Into Math after two years	-0.09 (0.07)	0.00 (0.01)	-0.03 (0.15)	-0.01 (0.03)	0.39** (0.11)	0.12* (0.05)
N students	13,448	13,448	12,762	12,762	13,429	13,429
N schools	37	37	37	37	37	37
N years	7	7	6	6	7	7

Source: School district data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses below the estimated coefficients. This table excludes early adopter schools that switched to Eureka Math in 2019–2020.

† Difference between Eureka Math and Into Math after the same number of years is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Composition of the comparison group

As discussed in Chapter III, there are different comparisons being averaged to estimate the effect of switching curricula. The comparison group in the main efficacy analysis comprises of a mix of four schools that did not switch curricula during the study period (but used other green curricula) and schools that had not yet switched to either Eureka Math or Into Math (most of which used a non-green curriculum before switching). This can make the main results more difficult to interpret, as the comparison condition reflects different counterfactuals. To understand how these comparisons are contributing to the estimates, we report findings when isolating the different comparison schools in Table B.4.

We restrict these analyses to two sets of comparison schools: (1) only the four schools that never switched curricula and (2) only the schools that eventually switched to the *same curriculum* in the future. To simplify comparisons to other results, we estimate effects only one year after the switch, separately by switch year and the curriculum that schools adopted. Note that to be able to compare only to schools that switched curricula in the future, these analyses focus on the effects for "early adopter" schools.

The results indicate that the estimated effects on NWEA MAP performance are not sensitive to the comparison group used for either Eureka Math or Into Math. Although the signs of the estimates are not stable when using these two restricted comparison groups, the magnitudes are generally small and are not statistically distinguishable from zero, as in the main analysis.

The estimated effects of switching to Eureka Math on state test performance are also not sensitive to the comparison group used, as the magnitudes are similar.[29] However, the estimated effects of switching to Into Math on state test performance do seem to be more sensitive to the comparison group composition. For Into Math, we find statistically significant negative effects on state test proficiency after one year of implementation among early adopters when compared to schools that switch to Into Math in the future, but small positive effects that are not significant when compared to schools that never switch.

---

[29] Note that due to the COVID-19 pandemic, results are not available for 2019–2020 switchers for STATE TEST scores.

When we examine math GPA and passing, we find that the large positive effects detected in the main analyses seem to be driven by the comparison to schools that never switched curricula. For both curricula, the estimated effects are small and occasionally negative when comparing to schools that would eventually switch to the same curriculum in the future. This suggests these findings may be driven more by the unique circumstances of schools that did not switch curricula during the study period.

Table B.4. Sensitivity of one-year impacts to different comparison groups of schools

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Schools that switched to Eureka Math in 2019–2020 vs. schools that never switched						
Eureka Math after one year	0.01	-0.00	NA	NA	0.25	0.09
	(0.03)	(0.02)			(0.17)	(0.07)
N students	7,519	7,519			7,486	7,486
N schools	23	23			23	23
N years	6	6			6	6
Schools that switched to Eureka Math in 2019–2020 vs. future Eureka Math switchers						
Eureka Math after one year	-0.04	0.00	NA	NA	-0.10	-0.02
	(0.04)	(0.01)			(0.11)	(0.04)
N students	10,413	10,413			10,321	10,321
N schools	36	36			36	36
N years	5	5			5	5
Schools that switched to Eureka Math in 2020–2021 vs. schools that never switched						
Eureka Math after one year	0.09	0.01	0.11	0.05	0.47**	0.14*
	(0.05)	(0.01)	(0.17)	(0.08)	(0.14)	(0.05)
N students	6,304	6,304	5,777	5,777	6,257	6,257
N schools	18	18	18	18	18	18
N years	6	6	5	5	6	6
Schools that switched to Eureka Math in 2020–2021 vs. future Eureka Math switchers						
Eureka Math after one year	0.03	0.01	0.08*	0.06	0.06	0.01
	(0.05)	(0.01)	(0.04)	(0.04)	(0.16)	(0.05)
N students	5,930	5,930	5,435	5,435	5,865	5,865
N schools	17	17	17	17	17	17
N years	6	6	5	5	6	6
Schools that switched to Eureka Math in 2021–2022 vs. schools that never switched						
Eureka Math after one year	0.05	0.01	-0.08	0.05	0.33	0.12
	(0.09)	(0.01)	(0.14)	(0.06)	(0.29)	(0.10)
N students	2,774	2,774	2,619	2,619	2,766	2,766
N schools	7	7	7	7	7	7
N years	7	7	6	6	7	7

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Schools that switched to Into Math in 2020–2021 vs. schools that never switched						
Into Math after one year	-0.01 (0.05)	0.01 (0.01)	0.06 (0.17)	0.02 (0.08)	0.42** (0.14)	0.14* (0.05)
N students	5,930	5,930	5,489	5,489	5,942	5,942
N schools	18	18	18	18	18	18
N years	6	6	5	5	6	6
Schools that switched to Into Math in 2020–2021 vs. future Into Math switchers						
Into Math after one year	-0.11 (0.08)	-0.01 (0.01)	-0.10 (0.05)	-0.12** (0.02)	-0.00 (0.22)	0.00 (0.10)
N students	4,992	4,992	4,618	4,618	5,007	5,007
N schools	16	16	16	16	16	16
N years	6	6	5	5	6	6
Schools that switched to Into Math in 2021–2022 vs. schools that never switched						
Into Math after one year	-0.06 (0.07)	0.01 (0.01)	0.16 (0.12)	0.04 (0.04)	-0.11 (0.29)	-0.01 (0.12)
N students	2,233	2,233	2,095	2,095	2,231	2,231
N schools	6	6	6	6	6	6
N years	7	7	6	6	7	7

Source: School district data.

Notes: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses below the estimated coefficients. The title of each panel indicates the schools included in each sample.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

NA = Not applicable. Because the state test was cancelled in 2019–2020, it was not possible to estimate the effects of switching to Eureka Math after one year for schools that switched in 2019–2020.

Appendix C

Follow-Up Analysis of Math Curriculum Efficacy Study

Introduction

Background

The Analysis of Middle School Math Systems (AMS) project is part of a larger set of investments by the Bill & Melinda Gates Foundation intended to support the math success of students who are Black, Latino, multilingual learners, or experiencing poverty. The study is organized around five broad inquiry areas (see box). This memo provides an update to a report from August 2023 focused on inquiry area 1: curricular efficacy. Based on a follow-up analysis that included an additional year of data, this memo updates findings on how the adoption of Eureka Math or Into Math by middle schools in one school district affected student math achievement.

Both focal curricula meet expectations for high-quality curricula, as defined by EdReports, and are therefore rated "green."[30] Following guidance from district leadership, between the 2019–2020 and 2021–2022 school years, most middle schools in the district switched to using either Eureka Math or Into Math. Of the 56 schools in the efficacy study, 36 schools (64 percent) adopted Eureka Math, and 16 schools (29 percent) adopted Into Math. Most of these schools switched from SpringBoard Mathematics, a "non-green" curriculum that does not meet expectations according to EdReports review criteria.[31] Finally, four schools (7 percent) did not switch curricula by 2021–2022. These schools continued using either Ready Mathematics or enVision Math, which are both rated green.

AMS study inquiry areas

4. Curricular efficacy
5. Curriculum characteristics that influence instructional enactment
6. Characteristics of professional learning that supports teacher needs and effective instructional enactment
7. Adaptations in instructional enactment
8. Factors that influence planned and unplanned adaptations in instructional enactment

The August 2023 report assessed the effects of these curriculum switches on (1) students' math performance on the Northwest Evaluation Association Measure of Academic Progress (NWEA MAP), (2) students' math performance on the state test, and (3) students' math course grades. We used a quasi-experimental difference-in-differences analysis to examine these effects up to three years after the curriculum switch occurred. This memo adds results from an additional school year, thus providing evidence on the effects of the curriculum switch up to four years after the switch occurred. With additional time, teachers might be able to overcome initial hurdles related to learning how to implement a new curriculum; thus, one possible hypothesis is that positive effects of switching to a stronger curriculum could emerge and strengthen over time as schools become more familiar with the material.

Overview of findings from 2023 report

Using district administrative data from 2016–2022, we found that switching curricula to Eureka Math or Into Math had no detectable effect on student achievement on standardized math tests but was

---

[30] Ratings were compiled by the authors from the publicly available EdReports Report Database.

[31] EdReports rates curricula as green—meets expectations; yellow—partially meets expectations; or red—does not meet expectations. These ratings are based on the degree to which materials are aligned with Common Core standards and meet their definition of usability.

associated with improvements in students’ math course performance in each of the three years of implementation we analyzed. The results were similar for both Eureka Math and Into Math. However, it was unclear whether the improvements in math course performance reflected increases in students’ math knowledge and skills. As we noted in the report, additional analyses suggested that the effects on course performance could have been due to other factors beyond increases in student learning, such as changes in teachers’ grading practices. Further, results from a Bayesian analysis indicated that the probability of positive effects for both curricula decreased with time after the schools adopted them.

We cautioned that these results should be viewed as being specific to the place and time in which the curricula were implemented. Our difference-in-differences design accounted for the overall impact of the COVID-19 pandemic across the district, and we did not find evidence that the effects differed by year schools switched curricula. Still, adopting a new curriculum during this period presented unique challenges that may have affected its effectiveness, even two to three years later.

Goals of the follow-up analysis

To better assess whether schools’ curriculum switch affected math achievement in the years after implementation, we obtained an additional year of administrative data from the school district for the 2022–2023 school year. This additional year of data allows us to measure outcomes up to four years after schools switched curricula, though only for schools that switched curricula in 2020 (which includes Eureka Math schools only). The additional year of data also allows us to update findings on the effects of switching curricula two and three years after implementation by including the schools that switched curricula in the later years of the study. For example, the earlier analysis of effects in the third year included only schools that switched curricula in 2020. Now, we can also include schools that switched curricula in 2021.

In this follow-up analysis, we also incorporated recent advances in the difference-in-differences literature. Recent research (for example, Goodman-Bacon, 2021) has cautioned that, when applied to instances when a new policy has a staggered adoption over time, the type of difference-in-differences estimation approach used in our 2023 report (known as a two-way fixed effects model) can introduce biases that compound in years further from implementation. In addition to reproducing the standard two-way fixed effects model with the additional year of data, we compared the results from the standard model to an alternative difference-in-differences estimator proposed by Callaway and Sant'Anna (2021). This comparison enables us to assess the risk of bias in the standard analysis given these recent advances.

Study design

Sample and data

The analysis in this memo used administrative records from the school district, which were available for all middle schools in the district and spanned multiple years. For our 2023 report, the school district provided student-level records for grades 6 to 8 from the 2015–2016 to the 2021–2022 school years. For this memo, the school district additionally provided student-level records for the 2022-2023 school year. All data included information on students’ background characteristics and math achievement, including their scores on the NWEA MAP and the state test and their letter grades in math courses.

Overview of analysis methods

Our analysis in this memo comprises three steps. First, we replicate the difference-in-differences analytic approach described in our 2023 report with the additional year of data to assess the effects of the curriculum change over a longer period of implementation. Second, we explore the sensitivity of these results to new methods being developed in the difference-in-differences literature. Finally, we apply a Bayesian interpretation to the updated findings.

The original approach from the 2023 report measures the effect of adopting a new curriculum by comparing (1) the change in trends over time for students at schools that switched to one of the two focal curricula with (2) the change in trends over time for students at schools that did not switch curricula (or had not yet switched). As discussed in our 2023 report, the difference-in-differences strategy can accurately estimate the effect of switching curricula, as long as these groups of schools showed similar average trends in their math outcomes in the years before schools switched (Angrist & Pischke, 2008; Imbens & Wooldridge, 2009). In addition, because adoption occurred on a staggered basis rather than all at once, the timing of adoption must not be related to differential trends in math outcomes (Baker et al., 2022).[32]

Because schools switched curricula over the course of a three-year period (2019–2020 to 2021–2022), the comparison group of schools used to estimate these effects can be a mix of schools that never switched curricula during this period and schools that had yet to switch at the time when the outcome was measured. For example, the estimated effect of switching to Eureka Math after one year is a weighted average of the following three types of comparisons:

1. The difference in trends between students at schools that switched to Eureka Math in the 2019–2020 school year and a comparison group of (1) schools that never switched, (2) schools that switched to Eureka Math or Into Math in 2020–2021, and (3) schools that switched to Eureka Math or Into Math in 2021–2022. This comparison used outcome data from the end of the 2019–2020 school year (the first implementation year for the 2019–2020 switchers).
2. The difference in trends between students at schools that switched to Eureka Math in the 2020–2021 school year and a comparison group of (1) schools that never switched and (2) schools that switched to Eureka Math or Into Math in 2021–2022. This comparison used outcome data from the end of the 2020–2021 school year (the first implementation year for the 2020–2021 switchers).
3. The difference in trends between students at schools that switched to Eureka Math in the 2021–2022 school year and students at schools that never switched. This comparison used outcome data from the end of the 2021–2022 school year (the first implementation year for the 2021–2022 switchers).

These comparisons can best be summarized as measuring the average effect of switching to Eureka Math or Into Math versus not switching, based on schools that never switched and schools that had not yet switched. Thus, to understand the comparisons being made, it is important to note which curricula schools switched from and which curricula schools that never switched continued to use. Most schools that adopted Eureka Math or Into Math (85 percent) switched from SpringBoard Mathematics, a non-

---

[32] Although "early adopter" schools had higher baseline math achievement than schools that switched curricula later, we found that their baseline *trends* were similar.

green curriculum.[33] The four schools that did not switch curricula continued using enVision Math (N = 1) or Ready Mathematics (N = 3), which are both rated green. [34] Therefore, the focal curricula are implicitly compared to a mix of non-green and green curricula.

Findings

Table C.1 presents the results of the main difference-in-differences analyses, which estimate the effects of switching curricula in the average district school after each year of implementation using the standard two-way fixed effects model. After incorporating an additional year of data into this analysis, the results remain generally consistent with those presented in the 2023 report.

As in the 2023 report, we found no evidence that switching to Eureka Math or Into Math improved students' math standardized test scores or the percentage of students who meet proficiency thresholds, on average, even four years after implementation.[35] In contrast, switching to a focal curriculum is associated with an increase in students' math GPA and passing rates in each year of implementation. When examining the effects of switching curricula separately for Eureka Math and Into Math, we again obtained similar results (Appendix Table C.1). The only positive effects we found are on course performance outcomes, with both Eureka Math and Into Math associated with increases in math GPA in the first two years. Additionally, switching to Eureka Math (but not Into Math) appears to increase the percentage of students with a passing math GPA.

Table C.1. Effects of switching to a focal curriculum, using two-way fixed effects

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
After one year	0.000	-0.5	0.085	4.7	0.354**	11.2*
	(0.050)	(0.6)	(0.113)	(3.5)	(0.097)	(4.7)
After two years	-0.035	-0.5	0.032	-0.5	0.349**	10.7*
	(0.063)	(0.5)	(0.096)	(2.6)	(0.090)	(4.6)
After three years	-0.086	-1.5**	-0.093	-5.3	0.292*	9.0
	(0.077)	(0.5)	(0.090)	(2.7)	(0.117)	(5.4)
After four years	-0.095	-1.6	-0.136	-8.2	0.495**	15.6*
	(0.094)	(0.8)	(0.115)	(4.6)	(0.170)	(7.1)
Number of students	22,170	22,170	21,246	21,246	22,118	22,118
Number of schools	56	56	56	56	56	56
Number of years	8	8	7	7	8	8

Source: School district data.

---

[33] Before switching to Eureka Math or Into Math, 28 schools had been using SpringBoard Mathematics, four schools had been using enVision Math, and one school had been using Ready Mathematics. Information on previous curriculum use was not available for 19 schools that switched curricula.

[34] Information on curriculum use was only available starting in 2019–2020, so we do not know which curricula schools used in the earlier years of the study. Thus, it possible that these schools switched curricula during an earlier period.

[35] Note that we can only measure a fourth year of implementation for schools that switched to Eureka Math.

Note: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. The effects after four years only reflect the effect of switching to Eureka Math. Standard errors are shown in parentheses.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

New developments in difference-in-differences methods

Limitations of the standard model for staggered adoption

As noted earlier, the estimated effects from the standard difference-in-differences model used in this study are a weighted average of the differences in outcomes between different groups. In this study, we examined four groups: schools that switched curricula in 2020, 2021, or 2022, and schools that did not switch during this period. In a "staggered adoption" such as this, the estimated effect of switching to a green curriculum is a weighted average of all possible differences between groups and periods.[36]

One major flaw of applying the standard difference-in-differences model to instances when there is staggered adoption has recently been highlighted in the literature. With a staggered adoption, the weights produced by a two-way fixed effects regression can vary wildly and are not motivated by an underlying rationale about which comparisons we may care about more (for example, Athey & Imbens, 2022; Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2021). In the worst case, the weights can even be negative. As a result, the estimates can be difficult to interpret and, especially in the case of negative weights, can be misleading.

Results using an alternative model

Recent advances in this literature have yielded alternative methods for estimating difference-in-differences designs that correct for the faulty weighting described above. One such method is that developed by Callaway and Sant'Anna (2021), which transparently allows the researcher to decide which comparisons are policy-relevant and use only those to estimate the desired effect. For example, we may be more interested in comparing schools that switched curricula to other schools that also switched curricula (but at a different point) than to schools that had already been using a green curriculum before the district issued the new guidance and therefore chose not to switch.

Table C.2 presents alternative estimates of the effect of switching to a green curriculum on NWEA MAP scores using the estimator developed by Callaway and Sant'Anna (2021).[37] The results show the effects of switching to a green curriculum using the following comparisons:[38]

1. Column (1) compares schools that switched to a green curriculum only to schools that eventually switched to a green curriculum but had not done so at the time the outcomes were measured. Given the comparison is only to schools that have not yet switched, it is not possible to measure effects over

---

[36] For example, one component of the final weighted average is the pre- versus pState test-switch difference in outcomes for schools that switched curricula in 2020 compared to the pre-pState test difference in outcomes for schools that switched in 2021. This difference is multiplied by a weight (a constant less than 1) and averaged with the pre- versus pState test-switch difference in outcomes for schools that switched curricula in 2020 compared to the pre-pState test difference for schools that switched in 2022, and so on (Goodman-Bacon, 2021).

[37] These are estimated using the recently developed "csdid" Stata package (Rios-Avila et al., 2023).

[38] We focus on just one outcome in the main text to facilitate discussion. See the appendix for additional results.

a longer implementation period, after all schools have already switched. Column (1) shows that the short-term effects of switching to a green curriculum on NWEA MAP scores were negative, on average. This effect seems to grow two years after implementation. Most previous studies of curriculum effectiveness measure progress over one academic year (Steiner, 2017); however, a prior Mathematica study that measured effects over time found estimates were attenuated after one year (Agodini et al., 2010).

2. Column (2) compares schools that switched to a green curriculum only to the four schools that did not switch curricula during the study period (these schools were already using a green curriculum). With this comparison group, it is possible to estimate the effects of switching curricula up to four years after implementation by comparing the schools that switched in 2020 to the four schools that did not switch. These findings suggest that switching curricula had no effect on NWEA MAP scores in the first two years after switching, and the effect became negative over time. However, this smaller group of schools may not offer an ideal comparison because they were already using a green curriculum before the new district guidance was announced.

3. Finally, column (3) includes both groups of schools from columns (1) and (2) in the comparison group. These findings are identical to the findings in column (2) for impacts after three and four years because these longer-run estimates rely solely on the comparison to schools that did not switch curricula during the study period. The impacts in column (3) after one and two years are a weighted average of the impacts from columns (1) and (2). Taken together, these findings suggest that the effect of switching curricula is small or non-existent in the short run and becomes negative over time.

   Column (3) is our preferred comparison because it uses the largest set of possible comparison schools in the earlier years, while allowing us to estimate longer-term effects. However, the same caveat applies as in column (2)—that we are including comparison schools that had already adopted a green curriculum, and this is the only available comparison for estimating the effects after three and four years. This was also true when using a two-way fixed effects estimator but was not as transparent.

Table C.2. Effects of switching to a focal curriculum, using different comparison groups with Callaway and Sant'Anna (2021) estimates

	NWEA MAP math z-score (mean)		
	Comparison to schools that switched in a later year	Comparison to schools that did not switch during study period	Comparison to both groups of schools
After one year	-0.082 (0.046)	0.024 (0.058)	-0.063 (0.049)
After two years	-0.138** (0.048)	0.085 (0.056)	0.085 (0.046)

	NWEA MAP math z-score (mean)		
	Comparison to schools that switched in a later year	Comparison to schools that did not switch during study period	Comparison to both groups of schools
After three years	n.a.	0.024 (0.109)	0.024 (0.109)
After four years	n.a.	-0.179* (0.087)	-0.179* (0.087)
Number of students	17,490	23,698	23,698
Number of schools	52	56	56
Number of years	6	8	8

Source: School district data.

Note: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. The effects after four years only reflect the effect of switching to Eureka Math. Standard errors are shown in parentheses.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

n.a. = not applicable. By the third year of implementation, there are no "not-yet-treated" comparison schools.

Reconciling the results

Using the methods developed by Callaway and Sant'Anna (2021), we can untangle two puzzles the results have presented based on the standard two-way fixed effects model. First, it was unclear why the effects appeared to grow increasingly negative over time, even for schools that switched curricula in the same year. Now it is clearer that as time goes on, the estimates rely more on the comparison to a small group of four schools that did not switch curricula during the study period. This is because they had already adopted another green curriculum before the school district issued the new guidance encouraging schools to adopt Eureka Math or Into Math. Policymakers and school leaders may not want to draw strong conclusions from this comparison group alone.

Another puzzling finding from the results of the two-way fixed effects model were the neutral or negative impacts of switching curricula on standardized test scores but large positive effects on math grades. Table C.3 presents the effects of switching curricula on math course performance separately for schools that switched to Eureka Math and Into Math, as well as for both groups combined, using the Callaway and Sant'Anna estimator and all available comparison schools (those that did not switch during the study and those that had not yet switched at the time of the comparison).

The findings show the positive effects of switching curricula on math course grades were present only among schools that switched to Into Math. These results make clear that the two-way fixed effects estimators were misleading—heavily weighting schools that switched to Into Math and possibly assigning negative weights to schools that switched to Eureka Math. Combined results with more meaningful weights show that the average impacts were negative and not significant after one, two, and three years, and thus more in line with the results on standardized test scores. Moreover, they make clear that the large, negative, and statistically significant estimates of the effects of switching after four years are completely dependent on comparisons to a small group of four schools and thus may not be illustrative of the broader impacts.

Table C.3. Effects of switching to a focal curriculum on math course grades using Callaway and Sant'Anna (2021) estimates, separately for Eureka Math and Into Math

	Eureka Math only		Into Math only		Any green curriculum	
	GPA (mean)	Proficient (%)	GPA (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
After one year	-0.107	-5.5	0.483*	20.6	-0.047	0.9
	(0.109)	(4.1)	(0.239)	(10.8)	(0.105)	(5.1)
After two years	-0.052	-2.9	0.359**	11.9*	0.014	-0.3
	(0.136)	(4.8)	(0.129)	(5.0)	(0.208)	(7.6)
After three years	-0.227*	-10.0*	0.047	1.8	-0.116	-6.3
	(0.097)	(3.9)	(0.122)	(5.4)	(0.107)	(4.6)
After four years	-0.456**	-22.4**	n.a.	n.a.	-0.456**	-22.4**
	(0.164)	(5.7)			(0.164)	(5.7)
Number of students	19,344	19,344	9,627	9,627	25,896	25,896
Number of schools	40	40	20	20	56	56
Number of years	8	8	8	8	8	8

Source: School district data.

Note: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

n.a. = not applicable. There are no schools that use Into Math through a fourth year of implementation.

Bayesian analyses

As in the 2023 report, we used a Bayesian interpretation to characterize the uncertainty of the estimated effects of switching to Eureka Math or Into Math (Deke et al., 2022). The estimated effects from the difference-in-difference analysis (either with two-way fixed effects or more modern estimators) can differ from the true effects for several reasons, including error in the measures of math achievement and variability in the sample. Using a Bayesian interpretation enables us to borrow information from other studies to try to reduce this uncertainty and estimate the probability that the effects were greater than different thresholds of interest.[39]

Given the caveats of the two-way fixed effects models outlined above, we focus on interpreting the Bayesian results for the Callaway and Sant'Anna estimates in the first two years after switching curricula, where the comparison group is the broadest (Appendix Table C.3).

---

[39] We used the BAyeSian Interpretation of Estimates (BASIE) tool from the Institute of Education Sciences to convert the main estimates into probabilities that switching to Eureka Math or Into Math had a positive effect. The BASIE tool enabled the study team to describe how likely it is that switching curricula in the study's context led to positive effects, given what is observed in this study's data and what previous studies examining middle school math outcomes have found.

Some evidence suggests that switching curricula may have helped students on the cusp reach the NWEA proficiency threshold, especially for those who switched to Into Math. This is despite limited evidence that switching curricula improved NWEA scores on average. Appendix Table C.3 reports higher probabilities of a positive effect the percentage of students who were proficient on the NWEA compared to a positive effect on average. Probabilities of a positive effect on the percent proficient in the first two years after switching to Eureka Math were between 78 and 86 percent and between 91 and 93 percent in the first two years after switching to Into Math. In comparison, the effects on average NWEA scores were likely to be positive 63 to 80 percent of the time.

The effects in the first two years of switching to a focal curriculum on math grades were higher for schools that switched to Into Math compared to schools that switched to Eureka Math. Consistent with the discussion of math grades above, the Bayesian probabilities reinforce the interpretation that switching to Into Math had a positive effect on math grades and the percentage of passing students in the first two years. The probabilities of a positive effect ranged from 95 to 100 percent. In contrast, the probability of a positive effect of switching to Eureka Math was low in the first year (27 percent for average GPA and 17 percent for passing rates)[40] and closer to neutral, or a coin toss, in the second year (between 37 and 48 percent) (Appendix Table C.3).

Conclusions

The latest results that incorporate an additional year of data, as well as recent methodological developments, suggest that the curriculum changes in the school district generally did not lead to sustained improvements in student math achievement. Although the COVID-19 pandemic presented unique challenges that likely made switching to a new curriculum even more difficult, some evidence indicates that the transition to Into Math may have been more effective than the transition to Eureka Math. The latest results based on newer methods suggest positive effects of switching to Into Math (but not Eureka Math) on math GPA after one and two years and possible negative effects of switching to Eureka Math on some outcomes (Appendix Tables C.2 and C.3). These results are consistent with survey data presented in the 2023 report. Specifically, that otherwise similar students had more positive beliefs about math when taught using Into Math than Eureka Math and that teachers in Into Math schools were more likely than teachers in Eureka Math schools to report that their curriculum was appropriate for their students.

This study highlights some of the difficulties in assessing the effects of curriculum adoption, especially when using quasi-experimental methods. Districts and schools make decisions about curricula on an ongoing basis, with the majority (57 percent) of U.S. school districts adopting a new curriculum every six to 10 years and an additional 14 percent replacing their curricula every one to five years (Allen & Seaman, 2017). Frequent changes in curricula, often implemented across entire districts, can make it difficult to identify appropriate, clear, and meaningful comparisons for study. Although in many cases quasi-experimental studies can yield similar results as experimental studies, the benefits of experimental studies may be especially useful for research on the effectiveness of curricula given these complexities.

---

[40] This is equivalent to saying the probability of a negative effect in the first year was 73 percent for average GPA and 83 percent for the passing rate.

Table C.4. Effects of switching to a focal curriculum using two-way fixed effects, separately for Eureka Math and Into Math

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Eureka Math after one year	0.029 †	-0.3	0.089	5.7	0.375**	11.8*
	(0.047)	(0.7)	(0.116)	(3.7)	(0.105)	(4.9)
Into Math after one year	-0.039	-0.2	0.100	3.4	0.323*	10.7
	(0.056)	(0.7)	(0.118)	(4.0)	(0.124)	(5.5)
Eureka Math after two years	-0.016	-0.0	0.052	-0.6	0.354**	11.2*
	(0.061)	(0.5)	(0.101)	(2.9)	(0.092)	(4.6)
Into Math after two years	-0.089	-1.1	0.005	-0.6	0.342**	9.8
	(0.073)	(0.9)	(0.096)	(2.5)	(0.122)	(5.8)
Eureka Math after three years	-0.082	-1.2*	-0.082	-5.5	0.312**	9.7
	(0.076)	(0.5)	(0.095)	(2.8)	(0.115)	(5.3)
Into Math after three years	-0.147	-2.8*	-0.142	-4.7	0.211	6.0
	(0.089)	(1.1)	(0.089)	(2.6)	(0.163)	(7.3)
Eureka Math after four years	-0.109	-1.9*	-0.147	-8.1	0.479**	14.9*
	(0.094)	(0.8)	(0.112)	(4.5)	(0.169)	(7.1)
Number of students	22,170	22,170	21,246	21,246	22,118	22,118
Number of schools	56	56	56	56	56	56
Number of years	8	8	7	7	8	8

Source: School district data.

Note: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses.

† Difference between Eureka Math and Into Math after the same number of years is statistically significant at the 0.05 level, two-tailed test.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

Table C.5. Effects of switching to a focal curriculum using Callaway and Sant'Anna (2021) estimates and all possible comparison schools, separately for Eureka Math and Into Math

	NWEA MAP math		State test math		Math course grades	
	z-score (mean)	Proficient (%)	z-score (mean)	Proficient (%)	GPA (mean)	Passing GPA (%)
Eureka Math after one year	0.021	0.7	n.a.	n.a.	-0.107	-5.5
	(0.061)	(0.7)			(0.109)	(4.1)
Into Math after one year	0.025	3.3	n.a.	n.a.	0.483*	20.6
	(0.161)	(2.2)			(0.239)	(10.8)
Eureka Math after two years	0.036	1.0	-0.287*	-15.3*	-0.052	-2.9
	(0.036)	(0.7)	(0.133)	(6.7)	(0.136)	(4.8)
Into Math after two years	0.014	1.3	n.a.	n.a.	0.359**	11.9*
	(0.044)	(0.8)			(0.129)	(5.0)
Eureka Math after three years	-0.017	0.5	-0.425**	-12.3**	-0.227*	-10.0*
	(0.098)	(0.5)	(0.088)	(3.8)	(0.097)	(3.9)
Into Math after three years	0.077	1.1**	n.a.	n.a.	0.047	1.8
	(0.108)	(0.4)			(0.122)	(5.4)
Eureka Math after four years	-0.179*	-1.7**	-0.188*	-10.5**	-0.456**	-22.4**
	(0.087)	(0.6)	(0.086)	(4.0)	(0.164)	(5.7)
Number of students	23,490	23,490	19,741	19,741	25,700	25,700
Number of schools	56	56	56	56	56	56
Number of years	8	8	7	7	8	8

Source: School district data.

Note: Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. Standard errors are shown in parentheses.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed test.

n.a. = not applicable. The school district's students did not take the state test in 2020 due to the COVID-19 pandemic. This missing year violates the csdid Stata command's requirement that the timing variable be measured in regular intervals. When calculating separate impacts for Eureka Math and Into Math schools, this missing year makes it impossible to estimate the one-year impact for Eureka Math and all impacts for Into Math on state test performance using the Callaway and Sant'Anna approach.

Table C.6. Bayesian interpretation of the estimates of switching to a focal curriculum from Callaway and Sant'Anna estimates

	NWEA MAP math				Math courses			
	Average z-score		Percent proficient		Average math GPA		Percent passing	
	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0	Estimated effect	Probability effect > 0
Eureka Math after one year	0.021	66%	0.7	78%	-0.107	27%	-5.5	17%
Into Math after one year	0.025	66%	3.3	93%	0.483*	96%	20.6	95%
Eureka Math after two years	0.036	80%	1.0	86%	-0.052	48%	-2.9	37%
Into Math after two years	0.014	63%	1.3	91%	0.359**	100%	11.9*	99%
Eureka Math after three years	-0.017	52%	0.5	74%	-0.227*	4%	-10.0*	2%
Into Math after three years	0.077	80%	1.1**	94%	0.047	70%	1.8	70%
Eureka Math after four years	-0.179*	7%	-1.7**	0%	-0.456**	3%	-22.4**	0%

Source: School district data.

Note: The estimated effects shown are the same as those in Table C.A.2. Standard errors are shown in parentheses. The probabilities shown reflect the likelihood that these effects were greater than the noted thresholds and were calculated using BASIE. Test scores were converted to z-scores by subtracting the mean and dividing by the standard deviation of scores for all students in that school year and grade level. As noted in Table C.A.2, the school district's students did not take the state test in 2020 due to the COVID-19 pandemic. This missing year violates the csdid Stata command's requirement that the timing variable be measured in regular intervals and makes it impossible to estimate the one-year impact for Eureka Math and all impacts for Into Math on state test performance using the Callaway and Sant'Anna approach. For this reason, we do not include the Bayesian interpretations of the state test estimates from the Callaway and Sant'Anna model.

* Coefficient is different from zero at the 0.05 level, two-tailed test.

** Coefficient is different from zero at the 0.01 level, two-tailed te

References

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., & Novak, T. (2010). *Achievement effects of four early elementary school math curricula* (NCEE 2009-4052). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Allen, I. E., & Seaman, J. (2017). *What we teach: K-12 school district curriculum adoption process, 2017*. Babson Survey Research Group.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Athey, S., & Imbens, G. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1), 62–79.

Baker, A., Larcker, D., & Wang, C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics, 144*(2), 370–395.

Callaway, B., & Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

Deke, J., Finucane, M., & Thal, D. (2022). *The BASIE (BAyeSian Interpretation of Estimates) framework for interpreting findings from impact evaluations: A practical guide for education researchers* (NCEE 2022-005). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/2022005/

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225(2),* 254-277.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 41*(1), 5–86.

Rios-Avila, F., Sant'Anna, P., & Callaway, B. (2023). CSDID: Stata module for the estimation of difference-in-difference models with multiple time periods. *Statistical Software Components,* S458976.

Steiner, D. (2017). Curriculum research: What we know and where we need to go. Standards Work. https://standardswork.org/wp-content/uploads/2017/03/sw-curriculum-research-report-fnl.pdf

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics,* 225(2), 175–199.

Mathematica Inc.

Our employee-owners work nationwide and around the world.

Find us at **mathematica.org** and **edi-global.com**.

Mathematica
Progress Together