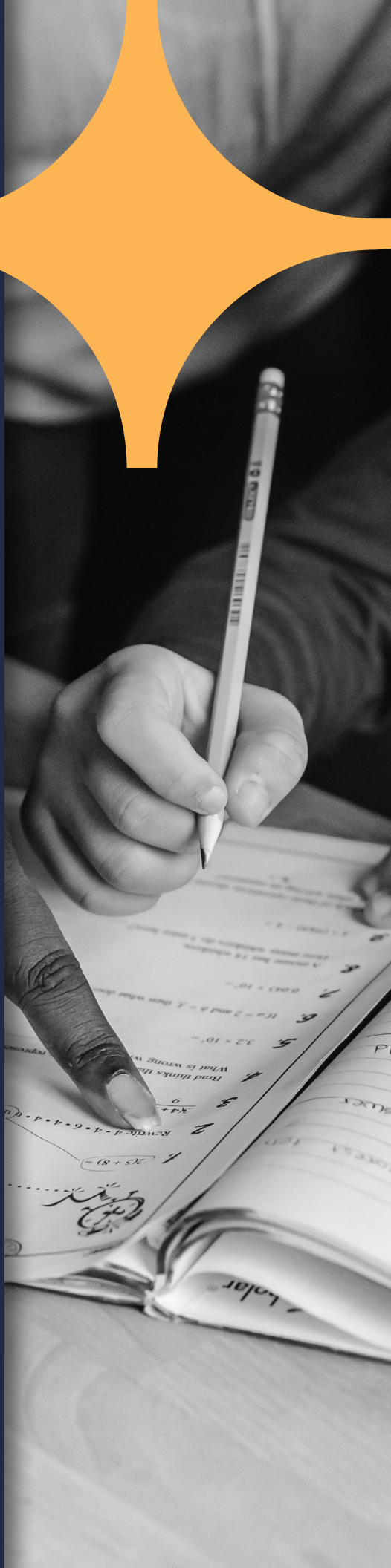
Bellwether

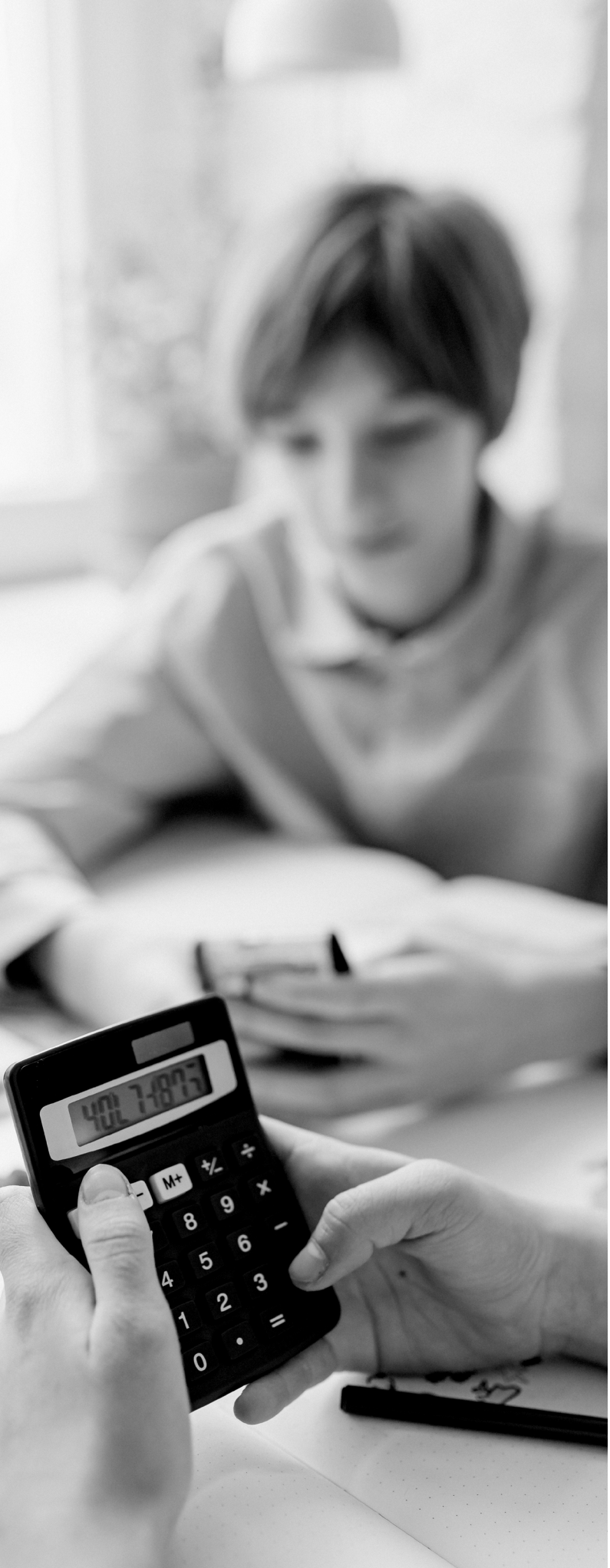
Rounding Up

An Analysis of Math Curriculum Effectiveness Studies

By Thomas Gold, Kristen Carroll, Leonard D.T. Newby, and Melissa Steel King

MARCH 2023





CONTENTS

Executive Summary

In 2022, the National Assessment of Educational Progress (NAEP) showed that 36% of fourth-graders and 26% of eighth-grade students across the United States were proficient in mathematics, representing a significant and alarming decrease at both grade levels since 2019. To address declining math proficiency across the country, educators are looking to adopt highly effective math curricula that rigorously meet state mathematics standards, improve teachers' content knowledge and instruction, and support the various needs of students.

Curriculum effectiveness research is an important source of information for educators looking to select a high-quality math curriculum that will meet their students' needs. Academic researchers, curricula publishers, and research institutions have produced a substantial body of research that examines the effectiveness of numerous math curricula. Yet, it remains unclear whether these robust studies address the factors that education decision-makers must consider when selecting among different math curricula.

In this analysis, we describe current approaches to assessing the effectiveness of math curricula and identify potential opportunities to further tailor effectiveness studies to the needs of the practitioners who rely on the research to make adoption decisions. Through analysis of 61 effectiveness studies, supplemented by expert interviews, we cataloged various characteristics (including the source, curriculum component of interest, unit of analysis, outcomes, research methods, duration of study, and others) and identified common themes and gaps.

Our analysis surfaced the following trends in our sample of math curriculum effectiveness studies:

- Student academic performance is the most studied outcome.
- The studies represent a mix of methodological approaches.
- Many of the studies used quasi-experimental designs (QED).
- There were few studies with experimental designs.

We also identified several limitations in this body of research:

- The studies were narrowly focused on student academic outcomes.
- Few studies examined teacher practice and perspectives.
- Few studies focused on program implementation.
- Few studies focused on student motivation.
- Few studies disaggregated results by student subgroups.

Based on these observed trends in current math curriculum effectiveness research, we developed the following recommendations for funders, researchers, and practitioners as they fund, design, and consume future research.

Funders

Fund more studies that use mixed-methods approaches.

Most effectiveness studies reviewed include only quantitative data, strongly leaning toward QED. While quantitative methods and research designs are important to understanding *if* a curriculum has an impact on student or teacher outcomes, it does very little to support our understanding of *how* the curriculum achieved various outcomes. For instance, was the curriculum effective because it was implemented with high fidelity? Did the curriculum materials guide the teachers in a new way, or was the math content simply more rigorous than previous curricula? By incorporating qualitative approaches within the research (e.g., teacher surveys, implementation observations), practitioners and researchers can better understand the context surrounding students' math growth.

Fund more studies that examine outcomes beyond student academic performance, such as teacher practices and student motivation.

Curriculum likely impacts students and teachers beyond assessment outcomes. For instance, curriculum may demonstrate success in improving teachers' pedagogical content knowledge or in improving students' attitudes toward learning mathematics. However, 72% of the studies reviewed defined "effectiveness" by measuring only student formative, embedded, or summative assessment outcomes. The field would benefit from more effectiveness studies that include outcome measures beyond student academic performance, helping practitioners understand how teachers and students respond to a new curriculum.

Fund studies that provide information about differential impacts on students with different characteristics.

Within the studies examined, only 31% disaggregated results by student subgroups (e.g., low-income, Black or Hispanic, English learner). Given the persistent gaps in student math achievement across subgroups and the pressures faced by many local education agencies (LEAs) to identify effective curricula for different student subgroups, future research should, to the extent possible, prioritize disaggregating the curriculum effects across student subgroups. Recognizing that academic researchers often have limited access to student-level data, curriculum publishers with close relationships to districts may be best positioned to conduct research that illustrates the differential impacts across student groups.

Researchers

Understand a program's theory of change when designing a math curriculum effectiveness study.

Not all curricula are designed to address the same topics, concepts, and mindsets; different programs present different approaches to learning math. Researchers should understand the theory of change behind any given curriculum. That understanding can inform how researchers determine appropriate measures and study curriculum design to assess its overall impact.

Include information about district context in the study.

Contextual factors contribute to curriculum success, and local district and school leaders must understand whether any given math curriculum will be a good fit. When curriculum effectiveness research incorporates information about contextual factors (e.g., level of teacher buy-in, training to support implementation, student access to tutoring) and how those factors may have contributed to the realized outcomes, education practitioners can better deduce if similar effects will likely occur within their context.

Practitioners

Consider more research-practice partnerships (RPP) or other similar arrangements to engage practitioners, communities, and policymakers in designing and implementing math curriculum effectiveness studies.

Collaboration between researchers and practitioners is an effective way to design relevant studies that can explore both short- and long-term outcomes for teachers and students. By engaging in RPPs, educators can have increased input in research designs that meet their local needs, and researchers can have opportunities to create long-term studies that lead to stronger insights on whether and how math curricula are effective. We encourage researchers and practitioners to engage in these partnerships and collaborate to understand what works in math curricula.

Consider opportunities to conduct rapid research.

While conducting a randomized control trial (RCT) experiment allows researchers to isolate the effects of a curriculum, RCT studies are often costly and time consuming for districts. To address this issue, we recommend that LEAs (potentially with researcher support) conduct smaller, rapid studies (three- to six-month pilots) of math curricula that can provide districts with the insights they need to expedite curriculum decisions.

Interactive Practitioner Planning Toolkit

For guidance on how practitioners can partner with curriculum providers, visit Bellwether's interactive planning toolkit.

Math Learning in the United States

Mathematics education and math learning in the U.S. are at a critical stage. Before the COVID-19 pandemic, there were early signs that math proficiency was improving nationwide, particularly for traditionally marginalized student populations.[1] Since 2020, pandemic-related factors and the shift to online learning have increased concerns about students' math proficiency. The 2022 NAEP math assessment showed that only 36% of fourth-graders and 26% of eighth-graders are proficient in math — a significant drop since the results of the same assessment in 2019.[2] This was the largest decline for NAEP math scores since the test was instituted in the 1970s, eradicating two decades of progress in math achievement.

These most recent NAEP results have broad implications for education in general, as research has shown that a lack of math proficiency in eighth grade is tied to lower performance in high school overall, including a decline in graduation rates.[3] In addition, the results raise concerns that fewer and fewer students in the U.S. are prepared to do advanced-level math, a factor that will have long-term negative effects on their success in higher education[4] and/or access to fast-growing economic sectors that require these skills in the workforce.[5]

To dramatically improve math learning, educators will need, in part, access to high-quality math curricula. However, understanding which curriculum is best for a specific school district can be challenging, as there are multiple factors to consider when choosing a program. Although there is a substantial body of research on math curriculum effectiveness, it is unclear whether these studies address all the factors that decision-makers need to provide schools with the best, most effective programs for their students. Such factors may include how the program is implemented, how well teachers are trained, or the degree to which the curriculum encourages students' motivation to learn.

Given the pressing need to provide students across the country with highly effective math instruction, this analysis examines the current state of math curriculum effectiveness research to gain insight into how researchers currently define and assess the quality of these materials and programs. We investigate trends in existing studies, with a focus on their methodologies and approaches to studying curriculum effectiveness. We also explore expert opinion on the state of the math curriculum research field.

The analysis concludes with a set of recommendations for funders, researchers, and practitioners. By conducting this analysis, we provide insight into areas of consensus around how to assess curriculum effectiveness and identify gaps in the literature. We also recommend promising potential approaches to studying math curricula. Ultimately, we identify opportunities to better provide school leaders with the information they need to make sense of the wide range of math curriculum research and select the right products for their students. Throughout the analysis, we provide snapshots of exemplar studies that highlight key findings in our analysis.

A Strong Math Foundation Is Critical to Student Success

The importance of a strong foundation in math is undeniable. Math achievement accounts for anywhere from 30% to 60% of the "variance in the chance of being on track to college readiness."[6] There is evidence that math achievement as early as pre-K predicts postsecondary outcomes. Overall, "school math achievement [is] a good predictor of whether students in [pre-K through grade 12] education stay on track toward two-year or four-year college education."[7]

Unfortunately, not all students have equal access to high-quality math instruction, a factor that may be contributing to persistent, substantive demographic differences in math proficiency scores across the country. On the 2019 NAEP mathematics assessment, on average white fourth-grade students scored 25 points higher than their Black peers and 18 points higher than their Hispanic peers.[8] These concerning disparities only increased during the pandemic; on the 2022 NAEP, the average mathematics score for white fourth-grade students was 29 points higher than their Black peers and 21 points higher than their Hispanic peers. Such results point toward growing and critical opportunity gaps in math education across the country, which only widen as students enter high school.

A critical component in the U.S. mathematics sequence is taking and passing algebra, ideally by the ninth grade. Students who take and pass Algebra 1 before ninth grade are more likely to go on to enroll in a postsecondary education program than students who take this course later in their high school career.[9] However, racial inequities exist in students' Algebra 1 course-taking before or in ninth grade. White, Hispanic, and Asian students are more likely to take Algebra 1 in ninth grade or earlier than Black students. In a recent study, 90% of white and 95% of Asian students took Algebra 1 before or in ninth grade, while 85% of Black students did the same.[10] The disproportionality in Algebra 1 course-taking in eighth grade can be attributed to both an access and an uptake issue; although about 80% of all eighth-graders attended a school that offered Algebra 1 in 2015-16, only about 24% of students across the country were actually enrolled in the course.[11]

Access to High-Quality Curriculum Is Critical to Improving Math Education

In general, access to a strong curriculum can influence student learning. One study showed that a shift to top quartile math textbooks increased student achievement by as much as .10 standard deviations, showing that "textbook choice is a high-stakes decision."[12*]

**We corrected the original version of this report, which incorrectly summarized the conclusion of the article cited in this sentence regarding the increase in student achievement associated with high-quality instructional materials. The article noted an increase in student achievement of up to .10 standard deviations, not an additional half-year of student learning.*

High-quality curricula are essential to helping students build a strong foundation in math. According to Dr. Matthew R. Larson, former president of the National Council of Teachers of Mathematics: "High-quality math content sparks student learning and engages students in investigating and discussing mathematics with the teacher and their classmates so that students develop a deep understanding of essential concepts."[13] In addition, high-quality math curricula should change the dynamic of learning from a traditionally passive approach, moving "students from simply watching the teacher to doing the math themselves," according to Larson.[14]

Determining the quality of a math curriculum requires understanding its degree of effectiveness on student learning. The National Academies of Sciences, Engineering, and Medicine defines math curriculum effectiveness as:

> ***"[Math curriculum effectiveness is] the extent to which a curricular program and its implementation produce positive and curricularly [sic] valid outcomes for students, in relation to multiple measures of students' mathematical proficiency, disaggregated by content strands and disaggregated by effects on subpopulations of students, and the extent to which these effects can be convincingly or causally attributed to the curricular intervention through evaluation studies using well-conceived research designs."*** [15]

A math curriculum is considered effective when different kinds of students who are taught with the curriculum demonstrate positive math results in multiple areas of math learning. Math achievement outcomes are one important indicator of curricular effectiveness; an effective math curriculum should also have a positive impact on other areas of student learning beyond test scores. For example, a high-quality math curriculum will improve students' mindsets toward the subject. By exposing students to engaging and motivating content, high-quality curriculum will improve their critical thinking skills and may encourage them to take higher-level courses later. In addition, these results should represent a continuation of positive performance or an improvement compared to a previous curriculum.[16]

To understand the impact of a particular curriculum, we rely on curricular effectiveness studies to examine the link between the use of the curriculum and various outcomes of interest. There are a variety of ways to approach research on curricular effectiveness. For more than a decade, the What Works Clearinghouse (WWC) set the standard for effectiveness research and is a trusted source for identifying effective curricula for all core subjects and more in schools. Teachers, administrators, researchers, and policymakers in education turn to the WWC for evidence of curriculum effectiveness on a variety of subjects as well as other measures of school effectiveness, such as teacher training, school culture, and character education programs. A product of the Institute of Education Sciences (IES) within the U.S. Department of Education, the WWC helps educational leaders make evidence-based decisions regarding curriculum using a consistent and transparent set of standards reviewed by more than 100 trained and certified reviewers.[17]

The WWC prioritizes a focus on causality — finding whether the curriculum or intervention in question causes certain student outcomes. As a result, the WWC reviews are frequently focused on QED or RCT because they are the most common approaches to determining causality. These approaches, which focus heavily on standardized student assessments, might not include a broader set of student outcome variables of interest that may be critical to many of the consumers of curriculum research. For example, school practitioners might want to also know about the implementation, which would necessitate a study that includes more qualitative research to help explain the phenomena.[18]

Experimental or quasi-experimental methods are not always the best or the only techniques for helping an LEA understand whether a particular curriculum is best for its students. When the goal of the research is to provide information to help educators select a product

— as opposed to, for example, understanding more about how students learn math, gaining insight into effective teaching practices, or validating a product for marketing purposes — it's not enough to simply assess whether a program causes certain academic outcomes in each sample of students. Rather, the practitioners who are primary consumers of math curricula (teachers, school leaders, and district leaders) also need to know which aspects of the program are relevant to *their* students and are viable to implement in *their* contexts. In the field of math curriculum effectiveness research, there is a clear need for studies focused on providing information about the fit and quality of the curriculum for the practitioners who will adopt, implement, and teach the materials.

This analysis explores the current landscape of math curricular effectiveness studies. We first provide an overview of the research methods we used to scan and summarize trends in the studies, through a combination of desk research and expert interviews. We then describe the characteristics of the 61 studies in our sample, identifying trends and gaps that we observed. Finally, we provide funders and researchers with recommendations for supporting and conducting future research on math curriculum.

This analysis provides researchers and funders with information on the current landscape of math curriculum effectiveness studies to contribute to the discussion of how research can be designed in a way to best support the needs of education practitioners and curriculum consumers. For researchers, we hope the information contained in this analysis will guide future studies in a way that their results will be most actionable to practitioners and policymakers as they make difficult decisions on curriculum adoption. We also hope philanthropists will find the information useful in their planning for future funding of research and evaluation studies with an eye toward building partnerships that support the interplay between research, practice, and evaluation.

Methodology

An Analysis of Current Effectiveness Studies

Given the sheer amount of effectiveness studies that exist in the field of math curriculum, we focused on research related to the most studied and widely used curricula in the field. We began this research by reviewing evaluation studies of math curricula rated by EdReports, a leading source for information about different K-12 curricula. EdReports is an independent nonprofit organization run by educators that provides assessments, including a quality rating, of instructional materials. There are currently around 215 different K-12 curricula reviewed on EdReports, with more than half of them (125) centered on math. EdReports rates each curriculum on two criteria:

1. **Alignment to college and career-ready standards**: Defined as "the degree to which materials meet expectations ... including that all standards are present and treated with the appropriate depth to support students in learning the skills and knowledge that they need to be ready for college and career."

2. **Usability**: Defined as the "degree to which materials are consistent with effective practices for use and design, teacher planning and learning, assessment, and differentiated instruction."[19]

EdReports uses a tiered system to rate curriculum alignment and usability, determining whether a certain curriculum meets expectations, partially meets expectations, or does not meet expectations on each dimension as determined by a set of review tools and processes. Reviews are conducted by educators.

Our review process to identify studies consisted of three stages:

1. **Identifying popular and/or well-researched curricula.**

To identify curricula, we filtered the 125 math curriculum reports on EdReports that met expectations for *both* alignment and usability, while also verifying that the curriculum was appropriate for grades K-8 and ninth-grade algebra, which yielded 46 different math curricula. We then widened our search criteria a little more during a second round by filtering for studies that met expectations for one of the criteria (alignment or usability) but earned only a "Partially Meets" ranking (or in some cases, "Not Rated") on the other. This allowed us to include an additional 13 math curricula, bringing our total up to 59.

We also reviewed the WWC database to see if there were any additional K-9 math curricula with evidence of effectiveness that were not included in the 59 identified through EdReports. Though we did not find any additional curricula, we were able to use the WWC to identify some additional studies associated with the 59 curricula.

2. **Identifying studies related to those curricula.**

We then went about finding studies that had been conducted on the 59 curricula we had identified. We did a rigorous search for research related to each curriculum on academic research databases, including Education Resources Information Center (ERIC), ProQuest, EBSCOhost, and Google Scholar. We used keywords ("curriculum name," "curriculum name and student outcomes," "curriculum name and teacher practice," "curriculum name and case study," "curriculum name and teacher practices," "math education research") and

searched across research databases including, but not limited to, ERIC, ProQuest, EBSCOhost, and Google Scholar. Additionally, we searched the websites for all 59 curricula for articles and reports published for potential customers. Of the 59 curricula identified as having a medium- or high-quality design by EdReports, the team found a total of 101 effectiveness studies highlighting 17 different curricula. It is not clear why the other 42 curricula lacked an associated study, though potential reasons include a lack of funding, time constraints, or a lack of cooperation with districts and other educational entities. In addition, there is the problem of publication bias; when no impacts are found, it's easy to stop and not publish the results. Finally, some curriculum publishers may invest in marketing research over conducting effectiveness studies.

3. Narrowing the sample down to relevant articles.

As we identified articles related to our target curricula, we reviewed each one for relevance by perusing their titles, abstracts, and methodology sections, among others. To be included in our review, the studies had to:

- Focus on K-9 math.
- Be conducted on math students in any U.S. state.
- Include a sample size of 250 participants or more.
- Publish in one of the following formats: peer-reviewed journal, final dissertation, curriculum publisher report, research report/white paper.

This process yielded a final sample of 61 different studies.

Once we finalized the sample of effectiveness studies, we cataloged the following features of each study:

Characteristics	Guiding Questions
Grade Level	What grades did the study focus on?
Source	Who conducted the study?
Curriculum Components of Interest	What aspects of curriculum were studied (e.g., just student outcomes or other components or product features)?
Outcomes	What outcomes were studied (e.g., effectiveness, rigor, cultural relevance, engagement, implementation conditions)?
Location	Where was the study conducted?
Sample Size and Population	What is the sample size (or N size) of the target population? Are there subgroups of students in the study?
Unit of Analysis	Does the study report outcomes by student, school, or another unit of analysis?
Research Methods	What are the different research methods that the study uses to collect and analyze data?
Findings	What are the study's findings about the effectiveness of the math curriculum?
Strengths/Limitations	What are the strengths and weaknesses of the research design?
Length of Study	How many years of data does the study examine? How long did the study take to finish?
Curriculum Format	What format is the curriculum delivered in?
Questions for Future Research	What, if any, questions for future research are raised by the authors? What is lacking in the study?

Interviews with Current Experts in Math Research

To gain additional context about the current state of math curriculum effectiveness research, we supplemented the literature review with interviews with seven math education experts, researchers, and/or curriculum publishers. Interviewees were identified and selected based on their connections to the field of research on math curriculum and math curriculum effectiveness. During the interviews, we probed for current thinking on methodological questions in math curriculum research such as:

- *How should we define "effectiveness" within the context of curriculum evaluations?*
- *What is the appropriate amount of rigor necessary to estimate the effectiveness of a math curriculum?*
- *What are the gaps and limitations in current math curriculum effectiveness studies?*

The interviews gave us an opportunity to understand the state of current math research through the eyes of leaders in the field. We were also able to explore the pros and cons of different approaches to study math curricula and identify ways that researchers can make their studies more relevant to practitioners and policymakers who are interested in not only finding high-quality curriculum but also in making an impact on math teaching and learning.

Limitations of the Analysis

There are some limitations to this analysis. First, our scan focused on effectiveness studies of high-quality curricula that were identified in EdReports and, as such, are not exhaustive. The studies of these other curricula, including those conducted outside of the U.S., may have offered additional insights and addressed some of the gaps identified in our review. Second, we limited our search to curricula for grades ranging from pre-K through grade 9 (with a focus on algebra for ninth grade), which eliminated several topics specific to high school mathematics in grades 10-12. Of the 59 curricula identified in EdReports, only 17 yielded research studies that met the strict requirements for our analysis. There are at least 42 math curricula evaluated by EdReports that schools are currently using that have not undergone this evaluation.

Finally, our analysis focused primarily on the research methods used in the 61 different studies and less on their actual findings (Appendix B).[20] For the purposes of this analysis, we were less interested in whether a specific curriculum is more effective than another, and more interested in understanding how the way the study is conducted — its research methods, the variables it measures, etc. — would affect its utility for a practitioner or another consumer of the research. That said, future explorations on this topic should look in more depth at outcomes to understand, for example, which specific curricula demonstrate more evidence of effectiveness with certain types of outcomes compared to others.

Characteristics of Current Math Curriculum Effectiveness Studies

General Characteristics

Overall, the 61 studies we reviewed for this analysis covered a broad range of characteristics, from the targeted grade levels to the types of research methods and outcomes studied. The studies we examined focused primarily on elementary-level math — 54% on curriculum that served grades pre-K to 5. An additional 5% of studies in our sample focused on the middle grades and 41% on a combination of grades. One focused on ninth-grade algebra as part of a combination of grades. Most of the studies focused on students without disaggregating results by student subgroups; only 31% looked at subgroups, such as students of different races and/or by gender.

The studies were conducted by a heterogeneous group of entities, split relatively evenly between academics and fee-for-service researchers. The plurality of the studies (34%) was conducted by an external (non-university-based) research group. In some, but not all cases, researchers were contracted by the curriculum publisher to study their product. Half of the studies were conducted by academic researchers — 25% by faculty researchers and another 25% by graduate students (dissertations) — and about 16% of the studies were conducted by the publishers themselves.

Most of the studies (54%) took place over the course of one year, 19% lasted three years or more, and 18% covered a two-year time period. There were also a small number of studies that were more rapid in nature, lasting less than a year.

As we describe in more detail below, most of the studies examined student test scores as outcome measures. Over half of the studies used standardized assessment as an outcome measure, with nearly 40% using summative assessments (including state tests) and 18% using one or more of the more common formative math assessments on the market (NWEA, etc.). Another 16% of the studies used assessments that were embedded in the curriculum itself. Only about 13% of the studies looked at teacher perspectives or practices as an outcome and 15% used a mixture of different outcomes. This finding reinforces the fact that researchers tend to treat curriculum efficacy studies as direct-to-student interventions without explicitly examining teacher effects, or the role that teacher or implementation characteristics play as mediators and moderators of efficacy.

Student Academic Performance Is the Most Common Outcome Studied

Of the 61 studies reviewed, 44 studies (72%) exclusively focused on the impact of the curriculum on student academic performance. Academic performance was measured by student math proficiency on formative assessments, summative state assessments, and curriculum-embedded assessments. Several of these evaluations (29 out of the 44) examined student academic performance outcomes that were aggregated at the group level (district, school, or classroom), rather than examining individual student-level data — likely due to the challenges associated with securing access to that data (Figures 1-6).

FIGURE 1: STUDY CHARACTERISTICS BY GRADE LEVEL

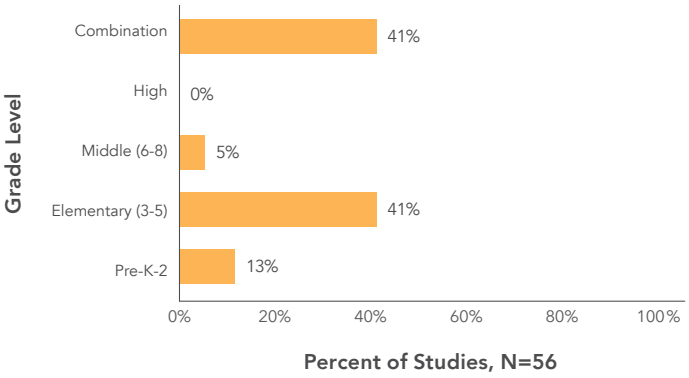


FIGURE 2: STUDY CHARACTERISTICS BY TARGET POPULATION

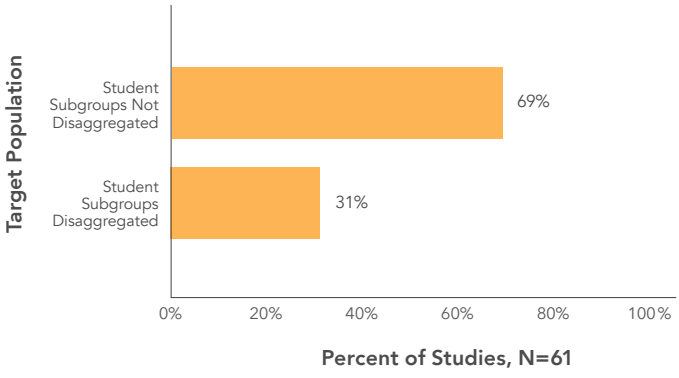


FIGURE 3: STUDY CHARACTERISTICS BY PERSON OR ORGANIZATION CONDUCTING THE RESEARCH

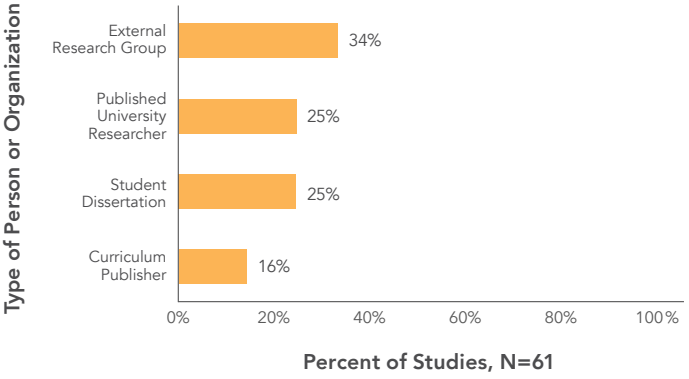


FIGURE 4: STUDY CHARACTERISTICS BY LENGTH OF STUDY

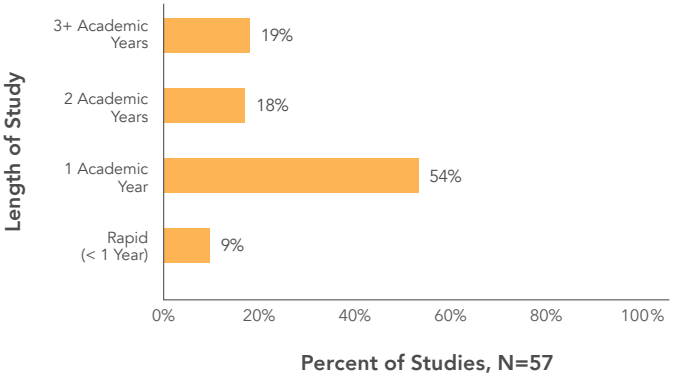


FIGURE 5: STUDY CHARACTERISTICS BY OUTCOMES STUDIED

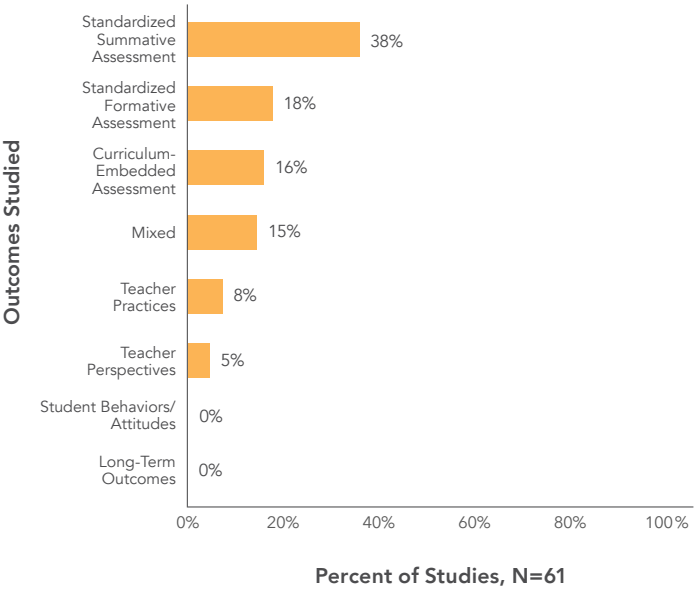
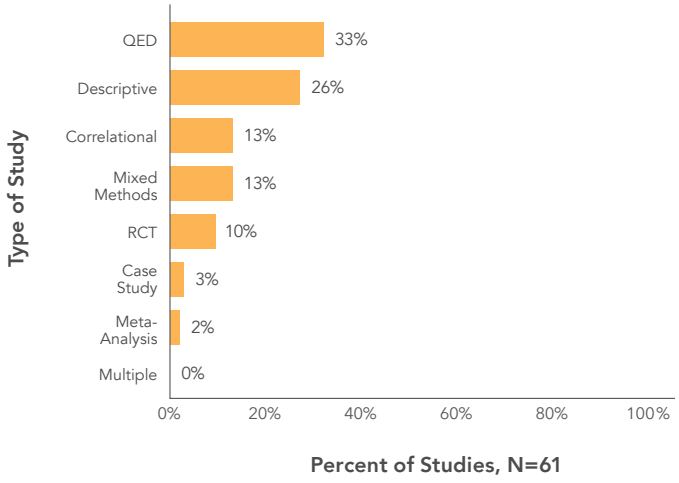


FIGURE 6: STUDY CHARACTERISTICS BY TYPE OF STUDY



***Note**: There are different N's for some of the study characteristics analyzed here due to missing values.*

Studies Tended to Use Quantitative Methods

Most studies in the analysis used only quantitative methods, whether in a descriptive, correlational, quasi-experimental, or experimental design as outlined below. Another 13% of the studies used a mixed-methods approach, which combined quantitative and qualitative approaches. Only a few of the studies collected qualitative data through observations and/or interviews and focus groups. Qualitative data can provide valuable information about the context of a new curriculum, including any challenges in its rollout, training the teachers, or getting buy-in from schools and key stakeholders. All these factors may be relevant in explaining curriculum outcomes.

Many Studies Used QED

One-third of the studies we examined implemented a QED. These studies used a rigorous method to create a synthetic comparison group, often through statistical techniques like propensity score matching — in which the comparison group is selected based on characteristics (like demographics and prior performance) that mirror those of the treatment group.

For example, an independent research organization completed a study of i-Ready, an online math curriculum, across the country of students in grades K-5.[21] The researchers first stratified their sample and matched students who were using a specific dosage of i-Ready in the 2018-19 school year to students with the same demographic characteristics who did not use i-Ready for mathematics in that time period. The study then used propensity score matching to pair students in the treatment and comparison group. The study used hierarchical linear modeling to estimate the impact of i-Ready on student achievement (measured by i-Ready student diagnostic assessments) and found that in each grade level, students who used i-Ready instruction for mathematics had a statistically significantly higher i-Ready diagnostic mathematics score.

The QED studies generally use extant data that is provided by schools or districts. The QEDs in the math effectiveness studies we examined often used state test scores or other existing outcome measures, either at the school or student level. QEDs meet the WWC's design standard "with reservations" and are often used to demonstrate some level of causal effects of an intervention, like a math curriculum. On the federal Every Student Succeeds Act's level of evidence pyramid, which was designed by IES/U.S. Department of Education, QEDs are at level 2, compared with RCTs at level 1. They are, however, more rigorous than correlational or descriptive approaches (levels 3 and 4, respectively).[22]

The QEDs in our sample reflect the use of rigorous designs in evaluation. This level of rigor is encouraged by districts and the federal government via policy mandates that require evidence of effectiveness before a curriculum can be purchased with public dollars. In addition, QEDs are typically much less costly than RCTs and easier to conduct in a real-world school setting, where it is often logistically difficult — and sometimes ethically questionable — to randomly assign one group of students to a particular curriculum or intervention.

Few Studies Used Experimental Designs

In effectiveness research, RCTs are often considered the "gold standard" for assessing the impact of a treatment or intervention because the design controls for any bias that may occur as a result of baseline variables.[23] In an RCT study, the "act of randomization … balances participant characteristics (both observed and unobserved) between the groups, allowing attribution of any differences in outcome to the study intervention."[24] This approach is considered the most statistically rigorous evidence of cause and effect between a particular aspect of a math curriculum and the outcome of interest (such as student math achievement).[25]

Ten percent of the studies that we examined successfully pulled off an experimental design — meaning an approach in which students are randomly assigned to the treatment or comparison groups. The small number of experiments is most likely due to the difficulty in implementing them, particularly in schools and school districts. Such studies involve considerable coordination to randomly assign participants to treatment and control groups and to build processes that reduce participant attrition.

Overall, while these studies provide important — and in the case of the RCTs and QEDs, rigorous — evidence about the effectiveness of math curriculum, there are still several gaps that present limitations in the utility of these studies. This is especially the case when we look at the capacity of these studies to inform instructional practices in schools. Some of these gaps include a strict focus on student achievement, a lack of focus on teacher practices and student motivation, very little focus on program implementation, and almost no focus on the district context. In the next section we cover each of these limitations.

Study Snapshot

One example of a math effectiveness RCT was conducted by the Rand Corporation in 2014. The study randomized schools into two groups over two years: one treatment group that received Cognitive Algebra 1 tutoring and one control group that relied on traditional algebra instruction. While first-year implementation effects were not statistically significant, second-year implementation at the high school level showed a statistically significant effect size of 0.2, a relatively small but significant result.

This finding indicates that two years into the implementation of this curriculum, student participants were significantly outperforming their peers in the comparison group. Questions for further research remained around the role of teachers, their level of curriculum adoption with fidelity, and their engagement over time.

Limitations in the Research

Most Studies Are Narrowly Focused on Student Academic Outcomes

Overall, using only state assessments to measure the effectiveness of a curriculum fails to capture the nuance of learning that takes place in a math class. Although math curricula are designed to impact student learning in multiple ways, student test scores and other similar outcomes do not always measure these critical components. While most high-quality curricula are aligned to the Common Core and/or state standards, they are also designed to develop specific sets of skills and dispositions, some of which may extend beyond the information tested on a state assessment.[26] Additionally, some researchers argue that standardized assessments do not measure the varied aspects of mathematical proficiency.[27]

Based on the framework "Adding it Up," mathematical proficiency includes students' demonstration of conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition.[28] Assessments of "disposition" occur via questionnaires or observations, a practice that is not typically included in state assessments.[29] As a result, research that observes only state assessment scores may miss key outcome measures of a curriculum's effectiveness intended by the curriculum developers in their theory of change.

In addition, by looking at broad academic performance measures, like scaled scores or proficiency levels, we may not be able to see how a curriculum can improve outcomes across various math domains like fractions or multiplication. Current evaluation studies that report the overall percentage of students meeting proficiency levels or the overall average scaled scores would provide few insights into how the curriculum impacted proficiency in specific domains. The need to evaluate the impact of a curriculum on specific domains is especially important in the subject area of math for early grades, since early content is foundational for later acquisition of higher-order math skills.

We recognize the accountability pressures U.S. educators face and the importance of understanding how a specific math curriculum will translate to state assessment scores. That said, a substantial body of research indicates that math curriculum can impact many aspects of student learning, and there is value in expanding effectiveness research to investigate whether specific curricula contribute to these other types of outcomes that educators care about.

Few Studies Focus on Teacher Practice and Teacher Perspectives

Although math curriculum users include teachers and students, we found that most of the curriculum effectiveness studies that we examined lacked a focus on the former. Before students receive a math lesson or are assessed on a state math test, teachers must engage with the curriculum to learn its content, understand the new ways of teaching math concepts, and then make decisions in the moment to assist in student learning. Because teachers are essentially the first users of a math curriculum, it is important to understand how they engage with the curriculum's materials, what they think about its quality, and how it may improve their overall practice. While curricula are designed to help students acquire math concepts and skills, it is important to remember that the curriculum "materials [aren't] doing the teaching ... the teacher is," according to Dr. Janine Remillard, professor and faculty director of teacher education at the University of Pennsylvania.

Even though teachers play an important role in math curriculum implementation, usage, and effectiveness, only eight of the 61 studies in our review solely examined teacher outcomes. Five of those eight studies examined teacher practice in response to curriculum implementation, and three explored teachers' perspectives throughout curriculum adoption. Another nine studies focused on mixed outcomes, which were a combination of teacher and student outcomes. Although these studies provided valuable information on how teachers engaged with the material, their findings were limited primarily because they used descriptive and mostly qualitative methods and focused on single cases of curriculum implementation, rather than examining multiple or comparative samples. This approach did not demonstrate a link between changes in teacher practice due to the curriculum and student outcomes.

Given the large body of literature providing evidence that teaching practice has a strong impact on student learning,[30] it is critical to understand how well a specific curriculum does or does not shape how math teachers approach instruction. Furthermore, as U.S. districts face increasing teacher shortages and educator pipeline challenges,[31] they are providing temporary or provisional teaching certificates to educators; LEAs have more teachers in the classroom who are not content specialists and may be less prepared. As a result, districts may be relying on curriculum materials to not only assist in improving student math knowledge and skill but also to improve teachers' conceptual knowledge of mathematics as well. Studies have shown that introducing a high-quality curriculum can help balance out the difference between novice teachers and more experienced ones.[32] Effectiveness research can therefore provide valuable information by examining how well different curricula support instruction[33] (e.g., by providing detailed scripts of teacher actions with a focus on math concepts[34]) and teachers' development of content knowledge.

In addition, new curriculum adoptions or changes often face pushback from teachers across districts and LEAs when their perspectives are not considered. While curriculum publishers understand the value of teacher perceptions and promote these views in

Study Snapshot

One exemplar of an effectiveness study that explored how curriculum influenced teacher practice was conducted by SRI and assessed Zearn Math Curriculum Study Professional Development.

Each teacher within the case study was provided Zearn professional development (PD) and explored each unit through small group lessons, collaborative analysis of student work, and a discussion of strategies for future problem-solving.

The study found that after a year of Zearn Math CS PD, teachers gained one level in their pedagogical content knowledge. The study also provided recommendations to Zearn on how to improve their PD, including "[differentiating] PD to meet the needs of a range of teachers."[35]

marketing materials, effectiveness researchers have not equally focused on teacher satisfaction or perceptions in general around math curricula. Among those who do consider teacher perceptions, the evaluations are primarily descriptive case studies, which are limited in generalizability for many practitioners. Overall, very few math curricula effectiveness studies in our sample examined how teachers interact with the curriculum, how a curriculum impacts teacher practice, or how changes in teachers' practice affect student outcomes.

Few Studies Focus on Program Implementation and Context

In our sample of curriculum effectiveness studies, most explored the relationship between curriculum adoption and student (and sometimes teacher) outcomes. But only 36% (22) of these studies explored the context of curriculum implementation and the role that this may have played in student outcomes. Several contextual factors could influence potential curriculum effectiveness. For example, districts vary in multiple ways, including the level of teacher autonomy in curriculum usage, the district's capacity to support teachers' adoption of a new curriculum, and the length of time districts invest in the rollout of a new curriculum before expecting results. Given this variation, effectiveness evaluations would benefit from including indicators of the quality of implementation of a new curriculum.[36]

One key characteristic of curricula implementation is the timeline of the rollout. For example, some curricula are implemented via pilot studies in particular grade levels and expanded each year, while others are implemented schoolwide in a single year. Among the 22 studies that examined curriculum implementation, only seven were more than one year in length. This finding is most likely a reflection of resource and time constraints on researchers as well as an expectation that change should be seen within one year of implementation. However, expert opinion[37] suggests that high-quality curriculum implementation is not achieved in only one year, but instead requires an ongoing, multiyear process of incorporating and adjusting the curriculum across

Study Snapshot

One exemplar evaluation study that successfully incorporated teacher curriculum engagement and student outcomes is an HMH Into Math study[38] focusing on K-8 teachers and students. To understand teacher engagement with the program, the evaluation collected data on teacher usage of the curriculum during the study period via teacher logs. The evaluation highlighted the reality that teachers engage with a curriculum to varying degrees and identified three different levels of fidelity (low, medium, and high) in teacher curriculum usage. In addition, the evaluators identified which materials within the curriculum were more engaging for teachers.

For example, the study found that teachers were more likely to use content focusing on "Building Understanding" and "Differentiated Options," rather than classroom activities such as the "Module Opener" task or "Turn and Talk." By combining information on teacher curricula usage and student outcomes, the study could test whether teachers' use of components of the curriculum was related to greater increases in student growth.

This kind of evaluation design helps practitioners decide how they might implement the curricula. It also enables practitioners to gain insight into what content within the curriculum is helpful to teachers. Furthermore, it provides information on the degree of curriculum usage that is needed to realize gains in student outcomes, something quite valuable to practitioners.

schools and classrooms. An effectiveness study may better capture the impact of a particular curriculum if it is conducted after the first year of implementation or spans multiple years (if budget allows), to enable teachers to adequately adopt the new curriculum and its concepts, and incorporate it evenly across their classrooms. Similarly, curriculum evaluations that help practitioners understand differences between implementation in year one and year three could provide key insights and set appropriate expectations for LEAs or practitioners considering new curricula adoption.

In addition to focusing on year one of curriculum implementation, most of the effectiveness studies in our sample did not provide much information about the fidelity of curriculum implementation within each district and/or school. Since teachers are often afforded a great deal of autonomy in many districts, evaluators may miss differences in the way the curriculum is used from classroom to classroom. While LEAs select a curriculum, teachers often have significant power to decide how that curriculum is used within their classroom. For instance, teachers may use a curriculum tool for most students within the classroom but use additional resources to supplement learning for other student subgroups who might be ahead or behind the level of the class. Similarly, it is common for teachers to apportion pieces of the curriculum with fidelity but use additional resources to cover other content. Without knowing if the curriculum was implemented to the degree its developers intended it to be, effectiveness research may be making broad assumptions regarding whether a particular curriculum is related to student outcomes or not.

It should be noted that technology-based math curriculum platforms may provide a unique opportunity to study implementation fidelity, as usage data is typically captured by the program software. However, active usage versus inactive time should also be distinguished if educators and researchers rely on online usage metrics to assess overall fidelity. For example, one Zearn Math study[39] compared consistent users to low users or nonusers in a coarsened exact matching analysis design. The study found an increase among all students who were consistently using the program.

Finally, there are numerous other factors that may influence the adoption of a new curriculum, including access to critical resources like coaches and trainers, the rate of teacher turnover, or even the political context around education. All these factors can influence the success of a curriculum rollout and eventual impact on students. As Dr. Trena Wilkerson, a professor at Baylor University, noted:

> ***"You have to know your student population, your context, your teachers, and their experience. What if you're in a school district where turnover is massive every year? If that's the case, then do you have a curriculum that stands on its own, [and can you] then support your teachers, new or not, coming in to be able to implement it with fidelity?"***

We understand that there is tension between study size and depth — it's rare a researcher can afford both. However, research that describes and evaluates variation in implementation can provide practitioners with the context needed to appropriately assess if a curriculum might work for them. While we recognize that it can be challenging for researchers to access the right information to assess program fidelity of implementation — such as metrics on usage and participation in training — without this context, researchers risk painting an incomplete picture of curriculum adoption and subsequent effectiveness. Whenever possible, including this information will help ensure that the findings are relevant and actionable for education practitioners and policymakers.

A final consideration related to implementation is that different LEAs often provide varying levels of support of PD for teachers when adopting a new curriculum. Among the studies in our sample, few examined the implementation of PD or its impact on curriculum effectiveness, despite its importance in setting the curriculum implementation up for success. Since PD equips teachers with tools to maximize student learning, its quantity, quality, and extent can strongly influence the effectiveness of a curriculum.[40] Thus, by analyzing

PD supports across curriculum users, an evaluation can provide additional insights on the curriculum's effectiveness.

Few Studies Focus on Student Motivation

Research shows that American students lag behind students from other countries in mathematics proficiency,[41] and that many students, particularly those from traditionally marginalized backgrounds, experience declining interest in and motivation to pursue the subject in postsecondary education.[42] A strong math curriculum can play an important role in the way students perceive math and can encourage a more positive, growth mindset toward the topic. Because student mindsets and beliefs about mathematics can be self-limiting,[43] it is important that effectiveness studies consider 1) how students vary in their motivation and how this interacts with their engagement with the curriculum, and 2) the degree to which curricula impact students' and teachers' mindsets and motivation toward math.

Unfortunately, while research on student mindset is growing, the effectiveness studies in our sample rarely considered students' math mindsets or their self-perceived mathematics abilities. Among the studies observed here, only one measured the impact of a particular curriculum on student perceptions or accounted for student perceptions within the study.[44] Practitioners would benefit from studies that examine students' math engagement and motivation, particularly when the curriculum they are exploring seeks to support the development of strong math mindsets (e.g., student perceptions and attitudes) as an outcome measure.

Study Snapshot

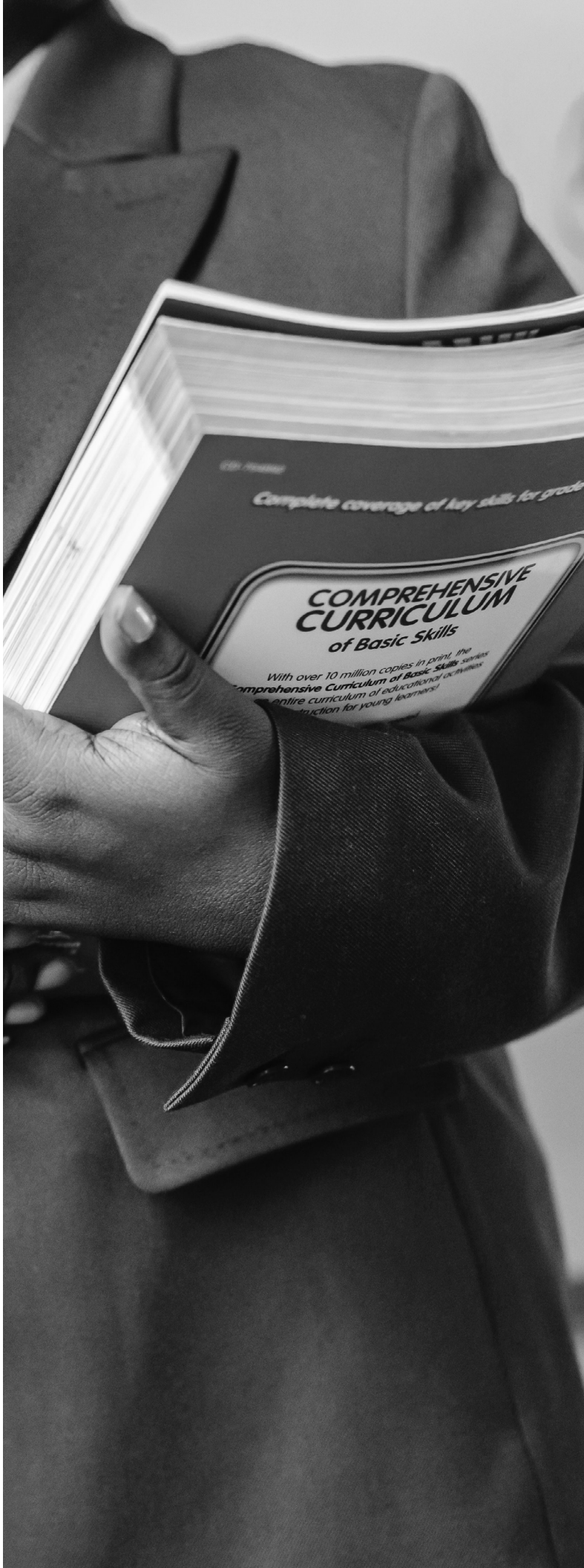
One example of a curriculum effectiveness study that incorporated the effects of PD in its analysis was an evaluation of the Stepping Stones Access program in multiple districts.[45] The study matched districts that purchased at least one day of professional learning and at least 10 days of professional learning with comparison schools that did not.

The study found a significantly positive difference in the average effect size for districts that purchased at least 10 days of professional learning but did not find statistically significant effects for districts that purchased at least one day. These results are valuable to districts, since they need to make decisions not only about which curriculum they should purchase, but also about how much support such as PD their teachers need to implement the curriculum successfully.

Few Studies Disaggregate Results by Student Subgroups

Many LEAs are seeking out curricula that will support their efforts to meet the needs of a diverse student population. Curriculum publishers have taken note and developed tools that provide various levels of differentiation to help teachers meet the needs of different learners within their classrooms. Yet, among the math curricula effectiveness studies examined in this analysis, we found that only 31% disaggregated results by various subgroups, including by race and/or gender. This has considerable implications not only for practitioners seeking to select a curriculum well suited to their particular student population, but also for research, which may be masking real areas of needs or strengths for specific groups of students.

One of the biggest challenges for researchers is getting access to data that would allow them to disaggregate results by race and other categories. Most of the studies we examined analyzed publicly available assessment data, often only at the school level. Access to individual student-level data is much harder to obtain, as it may require Institutional Review Board approval and, sometimes, the consent of a parent or guardian. These data access challenges may explain why two-thirds of the studies in our sample did not include information about the curriculum impact for different student subgroups. When possible, funders, publishers, and researchers should prioritize addressing these hurdles in order to disaggregate results, to increase the utility of the research for practitioners who may need to understand the effectiveness for a specific student subgroup before adopting the curriculum.



Recommendations

Findings from our review of math curriculum effectiveness studies point to a set of recommendations for funders, researchers, and practitioners. These recommendations identify opportunities for math curriculum effectiveness research to provide more actionable information that enables practitioners to identify high-quality curriculum that best meets students' needs.

Funders

Fund more studies that use mixed-methods approaches.

For social science research in general, mixed methods provide robust results.[46] Mixed methods in this analysis refer to studies that use a combination of both quantitative and qualitative data sources, including state administrative data (test scores, attendance, etc.) and surveys, along with qualitative methods like interviews, focus groups, and observations. By using both methods for collecting and analyzing data, studies ensure a clearer picture of not only the impact of a program but how it is implemented and why those results were achieved. According to Wilkerson, "The quantitative only tells you one piece, so you want to get more deeply into how that impacts individuals, students, and teachers that you're seeing."

Surveys and qualitative data are particularly important for understanding curriculum impacts. Qualitative information helps to provide insight into which component(s) of the curriculum is essential to achieving results — whether it's the instructional materials, the ways in which the materials guide the teacher, the students' or teacher's mindsets about the materials, etc. A mixed-methods study can identify these different aspects of a curriculum and describe their role in the impact on student outcomes.

Overall, though RCTs and QEDs provide one valuable source of information about program effectiveness, they do not necessarily present a holistic picture of how and why a given curriculum may influence student outcomes in each set of schools. While rigor is important to understanding any claims of causality, it cannot replace the need to understand the context of any student growth or change.

Fund more studies that examine outcomes beyond student academic performance, such as teacher practices and student motivation.

Effective curricula should have an impact not only on student academic outcomes, but also on how teachers teach the material (teacher practice) and how the material engages students' interest (student motivation). Yet most studies in this analysis assessed the impact of curricula on contemporary student academic performance, measured by state assessments.

Understanding what teachers think of a curriculum, how they engage with the materials, and how that impacts their instruction can provide valuable insight for districts looking to adopt curriculum that supports strong teaching practices. Researchers should also assess how students engage with a curriculum and if their attitudes and interest in mathematics changes because of the curriculum adoption. Together, these additional outcomes help expand upon why certain curricula are (or are not) effective.

Funders should consider investing in future math curricula effectiveness research that includes outcomes such as:

- Student mathematics motivation and attitudes.
- Long-term student outcomes (e.g., future math interest, course selection, and college readiness).
- Teacher practices in the classroom.
- Teacher pedagogical content knowledge.
- Teacher perceptions and/or beliefs around a specific curriculum.

Fund studies that provide information about differential impacts on students with different characteristics.

The research in our sample rarely disaggregated results by student subgroups. Given that most researchers used group-level student assessment data, it is likely that disaggregated subgroup data was often unavailable. Before investing in or adopting a new curriculum, many LEAs may want to know the impact of a curriculum across various student groups in terms of student race, household income, and whether they are English learners. If there are barriers to understanding the impact of the curriculum at this level, then researchers should partner with districts to assess if curricula demonstrate various impacts across student subgroups.

In general, researchers who do not have more formal relationships with districts may find it hard to access the student-level data needed to conduct subgroup analysis. Curriculum publishers, who often have close relationships with districts, may be in a better position to collect and analyze student results at the subgroup level, though there might be questions around objectivity. Researchers can also consider creating RPPs with districts to gather more granular data for such analyses, as outlined below.

Researchers

Understand a program's theory of change when designing a math curriculum effectiveness study.

Not all curricula are the same and, ostensibly, every new curricular program presents a different, perhaps novel, approach to learning math. It is only by understanding the way the program is intended to teach math that appropriate outcome measures can be selected. As Dr. John Pane, senior scientist from the Rand Corporation, pointed out, "A theory of change is a valuable guide to measurement and study design." For example, if a program seeks to change students' motivation around math as a catalyst for deeper learning, then outcome measures should move beyond summative academic assessments and include assessments of grit and self-efficacy. In addition, the approach of a particular math curriculum may also dictate the length of time it should be implemented before being studied.

Researchers should use information drawn from the theories of change to create the study design and develop effective measures that accurately assess all aspects of a program. These measures may include student or teacher surveys to understand the impact of the program in nonacademic ways. Using a mixed-methods approach, researchers may also include more classroom observations to see how the curriculum may change the dynamic between students and teachers.

Include information about district context in the study.

For most districts, curriculum adoption includes a variety of factors, from access to training and other resources, to local education politics and teacher buy-in. Yet, current curriculum studies rarely address the context surrounding curriculum adoption or implementation. While these factors may seem negligible within a research design, they provide practitioners with information that may strongly influence their decision to adopt a curriculum or impact their likelihood of successful implementation. For example, adoption of a new curriculum may go more smoothly in a district with a high ratio of school-level math coaches and low teacher turnover. Such a district may see more positive results than a district that lacks human capital and other resources. Similarly, studies[47] of curriculum implementation identified that not only are there various approaches to implementation, but there are also a variety of potential teacher-centered and student-centered problems that a district may encounter while implementing a curriculum.[48] Because contextual factors do contribute to curricular successes, more research should devote time to understanding and measuring these factors. Potential measures of district context include, but are not limited to:

- The type and degree of district support that is available during a new curriculum implementation.
- The overall capacity that is available to support sustained PD (i.e., math coaches at every school versus one math coach for the district).
- The degree of teacher buy-in during curriculum implementation.
- The state or local pressures around student achievement facing districts when adopting a new curriculum.

Practitioners

Consider more RPPs or other similar arrangements to engage practitioners, communities, and policymakers in designing and implementing math curriculum effectiveness studies.

As collaborations among researchers, districts, and communities, RPPs produce research that is rigorous in methods and relevant in output.[49] Through these arrangements, the collaboration starts at the beginning — developing the research agenda — and continues through the research process. RPPs provide opportunities for greater input from key stakeholders, paving the way for more relevant research outputs that are aligned with local needs. RPPs are also not monolithic and have been designed in multiple ways to incorporate local issues.

By engaging in RPP-type practices, researchers can work more closely with practitioners in long-term studies that will provide increased access to data on district context and implementation, explore short- and long-term outcomes for teachers and students, and create stronger insights on what, how, and why math curricula are effective or not. Such partnerships are also effective at giving researchers access to highly sensitive data, like individual student test scores, which allow them to conduct the rigorous and robust research that is often needed to understand what works in math curricula.

Consider opportunities to conduct rapid research.

Although there may be concerns that studies under one year in length might not elicit results that are robust enough for informing decisions, studies that are too long risk not being relevant to the procurement process. It is a best practice that LEAs pilot new curricula before a full rollout, though conducting an RCT is a long and expensive endeavor. While RCTs assist in rigorously identifying the causal effects of a particular curriculum, sometimes the results come far too late for schools and districts looking to decide between curricula. This points to an overall efficiency problem — there is a lot of math curricula that needs to be evaluated, and RCTs are not a very efficient way of doing it. Even supporters of RCTs recognize this challenge. According to Pane, "Because RCTs are so expensive and time-consuming and the number of products out there is so big, it's impossible to have evidence on all of them, at least rigorous evidence."

To address this issue, districts should consider conducting smaller, more rapid studies that can help them efficiently assess whether a particular program is the right one for their students. Some savvy LEAs develop their own research teams to assess effectiveness of a particular curriculum within their district. In addition, RPPs, researchers, and funders can bridge the gap to meet districts' needs by providing them with three- to six-month pilot studies to guide decision-making and build long-term partnerships.

Conclusion

Along with the need to improve teaching and learning of math, there is a commensurate need to build an evidence base for high-quality K-12 math curricula in the U.S. that practitioners can use to make effective decisions. A robust analysis of existing studies, including everything from strict quantitative examinations of student outcomes to nuanced mixed-methods approaches, amplifies that need.

Many studies included in this analysis provide valuable information on the effectiveness of different curricula, but gaps remain. Few studies focus on factors like the institutional context of the implementation of new curriculum, teachers' perspectives, student motivation, or even different characteristics of students. Funders, researchers, and practitioners must address these gaps (though not necessarily all in the same study) to make informed decisions on high-quality math curriculum adoption that can improve math instruction and learning for K-12 students across the country.

For guidance on how practitioners can partner with curriculum providers, visit Bellwether's interactive planning toolkit. ✦

Appendix A. Studies Reviewed

Curriculum	Grade Level	Study
Bridges in Mathematics	Elementary (3-5)	1. Garret J. Hall, Patti Schaefer, Teri Hedges, and Eric Grodsky, "Examining Bridges in Mathematics and Differential Effects Among English Language Learners," *School Psychology Review* 51, no. 4 (2022): 392–405.
Carnegie Learning Middle School Math	Middle (6-8)	2. Carla Simmons White, "A Study of the Implementation of a Middle School Math Program and Student Achievement," Gardner-Webb University, 2018.
CK-12 Interactive Math, Singapore Math, and Prentice Hall	Combination	3. Leanna R. Carollo, "Beyond Elementary: Examining Conceptual Demands of Division of Fractions in Current U.S. Curricula," *Oregon Undergraduate Research Journal* 4, no, 1 (2013): 35–53.
Cognitive Tutor Algebra 1	Combination	4. John F. Pane, Beth Ann Griffin, Daniel F. McCaffrey, and Rita Karam, "Effectiveness of Cognitive Tutor Algebra I at Scale," *Educational Evaluation and Policy Analysis* 36, no. 2: (2016): 127–144.
Connected Mathematics	Middle (6-8)	5. Julie E. Riordan and Pendred E. Noyce, "The Impact of Two Standards-Based Mathematics Curricula on Student Achievement in Massachusetts," *Journal for Research in Mathematics Education* 32, no. 4 (2001): 368–398.
Discovering Mathematics: Algebra, Geometry, Advanced Algebra	Combination	6. Agile Mind Intensified Algebra I course.
EnVision	Combination	7. Miriam Resendez and Mariam Azin, *A Study on the Effects of Pearson's 2009 enVisionMATH Program*, 2007-2008: first year report, PRES Associates, Inc., 2009. 8. Cory Koedel, Diyi Li, Morgan S. Polikoff, Tenice Hardaway, and Stephani L. Wrabel, "Mathematics Curriculum Effects on Student Achievement in California," *Aera Open* 3, no. 1 (2017): 1–22. 9. David Blazar et al., "Curriculum Reform in the Common Core Era: Evaluating Elementary Math Textbooks Across Six U.S. States," *Journal of Policy Analysis and Management* 39, no. 4 (2020): 966–1019. 10. Roberto Agodini, Barbara Harris, Neil Seftor, Janine Remillard, and Melissa Thomas, "After Two Years, Three Elementary Math Curricula Outperformed a Fourth," NCEE evaluation brief, September 2013. 11. Miriam Resendez and Mariam Azin, *A Study on the Relationship Between Pearson's 2009 enVisionMATH Program and Student Math Performance Among English Language Learners, Minorities, and Economically Disadvantaged Students: Special Report* (Jackson, WY: PRES Associates, Inc., 2010). 12. "enVisionmath2.0 Research Overview. Evidence of Effectiveness: A Summary of the 2015-2017 Longitudinal Efficacy Study," Savvas Learning Company, 2020, https://assets.savvas.com/asset_mgr/current/202038/enVm2_RCT_Research_Overview_Savvas.pdf.

Appendix A continues from page 27-32.

Curriculum	Grade Level	Study
Eureka Math	Elementary (3-5)	13. Amy Dawn Dwiggins, "Educative Features of Upper Elementary Eureka Math Curriculum," doctoral dissertation, University of Missouri-Columbia, 2020. 14. Ticada Guyton, "Correlation of Teacher Perceptions of the Expeditionary Learning and Eureka Math Curricula and Student Achievement," doctoral dissertation, Union University, 2021. 15. Melissa M. Lein Authement, "A Case Study of Four Teachers' Experiences While Implementing the Latest Version of the Eureka Math Curriculum in the State of Louisiana," doctoral dissertation, Baylor University, 2022. 16. Nicole Whitehurst, "An Analysis of Eureka Math Curriculum for Common Core Alignment and Development of Conceptual Understanding," doctoral dissertation, University of Oregon, 2016. 17. Tiah B. Alphonso, "Investigating Curriculum Use and Its Impact on Teachers and Their Practice," doctoral dissertation, Louisiana State University, 2016.
Everyday Mathematics	Pre-K-2	18. Karen C. Fuson, William M. Caroll, and Jane V. Drueck, "Achievement Results for Second and Third Graders Using the Standards-Based Curriculum Everyday Mathematics," *Journal for Research in Mathematics Education* 31, no. 3 (200): 277–295. 19. Mariann T. Helfant, "The Relationship Between Third and Fourth Grade Everyday Mathematics Assessments and Performance on the New Jersey Assessment of Skills and Knowledge in Fourth Grade (NJASK/4)," Seton Hall University, 2005. 20. Julie E. Riordan and Pendred E. Noyce, "The Impact of Two Standards-Based Mathematics Curricula on Student Achievement in Massachusetts," *Journal for Research in Mathematics Education* 32, no. 4 (2001): 368–398. 21. William M. Ward, "The Influence of a Reform-Based Mathematics Program on Third, Fourth and Fifth Grade Student Achievement," Seton Hall University, 2009. 22. Andy Isaacs, Traci Higgins, and Catherine Randall Kelso, "The ARC Center Tri-State Student Achievement Study," National Science Foundation, 2003. 23. Vanessa Constance Beauchaine, "Differentiating Instruction to Close the Achievement Gap for Special Education Students Using Everyday Math," Boston College, 2009.

Appendix A continues from page 27-32.

Curriculum	Grade Level	Study
HMH into Math	Combination	24. JEM & R LLC, "Research Results: HMH Into Math, K-8 Early Outcomes 2020SY," Houghton Mifflin Harcourt, 2020, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HMH-Into-Math-Research-Results-Paper-2019-2020-SY.pdf. 25. JEM & R LLC, "Research Results: HMH Into Math Full-Year Impact Study 2021SY," Houghton Mifflin Harcourt, 2022, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HMH-Into-Math-Research-Results-2021SY-FINAL-06-07-22.pdf. 26. Laura Davis and Candy Jones Martorell, "Effects of Implementing One Math Curriculum from Kindergarten through 12th Grade on Teacher and Student Efficacy and Achievement," doctoral dissertation, Lipscomb University, 2018, https://www.proquest.com/openview/443f4ff746495591ac34b36a5e30e9d9/1?pq-origsite=gscholar&cbl=18750.
HMH Into Math Florida	Combination	27. "HMH Into Math 2020: Grade 5 Florida Pilot," Educational Research Institute of America, 2019, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HMH-Into-Math-FL-Spring-2019-FT.pdf.
i-Ready Classroom Mathematics	Combination	28. Matt Dawson, *The Impact of COVID-19 on Student Academic Growth in 2020-2021*, report no. 19, Curriculum Associates, December 2021, https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-covid-growth-research-paper-2021.pdf. 29. "Impact of Ready Reading and Ready Mathematics on Student Learning," research brief, Curriculum Associates, January 2022, https://www.curriculumassociates.com/-/media/mainsite/files/ready/ready-essa-brochure.pdf. 30. Tracy L. Lewis, "i-Ready Mathematics Effectiveness on Student Achievement and Teacher Evaluation Scores: A Quantitative Study," dissertation, University of Phoenix, 2018. 31. Matthew Swain, Bruce Randel, and Rebecca Norman Dvorak, "Impact Evaluation of Mathematics *i-Ready Instruction* for Elementary Grades Using 2018-19 Data: Final Report," no. 106, Human Resources Research Organization (HumRRO) for Curriculum Associates, 2020. 32. Chelshea Pruznak, "The Effectiveness of i-Ready Instruction on Student Growth," doctoral dissertation, Indiana University of Pennsylvania, 2021. 33. Bruce Randel, Matthew Swain, Rebecca Norman Dvorak, Elisabeth Spratto, and Jordan Yee Prendez, "Impact Evaluation of Mathematics 'i-Ready' for Striving Learners Using 2018-19 Data: Final Report," no. 048, online submission, 2020. 34. Michael A. Cook, Ashley A. Grant, and Steven M. Ross, "The Impacts of i-Ready Personalized Instruction on Student Math Achievement in Multiple School Districts," Center for Research and Reform in Education, 2022.

Appendix A continues from page 27-32.

Curriculum	Grade Level	Study
i-Ready Classroom Mathematics *cont.*	Combination	35. Matthew Swain, Bruce Randel, and Rebecca Norman Dvorak, "Impact Evaluation of Reading 'i-Ready Instruction' for Elementary Grades Using 2018-19 Data: Final Report," no. 107, Human Resources Research Organization, 2020. 36. Madison A. Holzman and Molly K. Duncan, "Impact of *i-Ready Personalized Instruction* on the Massachusetts Comprehensive Assessment System Achievement for Grade 5 in English Language Arts and Mathematics," report no. 07, Curriculum Associates, 2022, https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-impact-on-mcas-report-2022.pdf. 37. "Evidence of the Impact of *i-Ready* on Students' Mathematics and Reading Achievement," Curriculum Associates, 2019, https://www.sbsd.k12.ca.us/cms/lib/CA01001886/Centricity/Domain/41/iready-research-efficacy-brief-2019.pdf. 38. Ricardo Alberto Torres, "The Effect of the i-Ready Reading Program on Student Scores on the Northwest Evaluation Association (NWEA®) Measures of Academic Progress (MAP) Reading Assessment," dissertation, Cleveland State University, 2019. 39. Kenyatta Shanta Aldridge, "The Relationship Between I-Ready Intervention and Grade 8 Mathematical Achievement," doctoral dissertation, Columbus State University, 2021, https://csuepress.columbusstate.edu/cgi/viewcontent.cgi?article=1438&context=theses_dissertations.
Math Expressions	Pre-K-2	40. Roberto Agodini, Barbara Harris, Melissa Thomas, Robert Murphy, and Lawrence Gallagher, "Achievement Effects of Four Early Elementary School Math Curricula: Findings for First and Second Graders" (NCEE 2011-4001), National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010, https://files.eric.ed.gov/fulltext/ED512551.pdf. 41. Roberto Agodini, Barbara Harris, Sally Atkins-Burnett, Sheila Heaviside, Timothy Novak, and Robert Murphy, "Achievement Effects of Four Early Elementary School Math Curricula: Findings From First Graders in 39 Schools" (NCEE 2009-4052), National Center for Education Evaluation and Regional Assistance, 2009. 42. Roberto Agodini and Barbara Harris, "How Teacher and Classroom Characteristics Moderate the Effects of Four Elementary Math Curricula," *The Elementary School Journal* 117, no. 2 (2016): 216–236. 43. Roberto Agodini, Barbara Harris, Neil Seftor, Janine Remillard, and Melissa Thomas, "After Two Years, Three Elementary Math Curricula Outperformed a Fourth," NCEE evaluation brief, September 2013. 44. "Success Across Wisconsin: Case Studies From Selected Wisconsin School Districts Using Houghton Mifflin Harcourt Math Expressions," Houghton Mifflin Harcourt, 2013, http://hmhco-v1.prod.webpr.hmhco.com/~/media/sites/home/educators/education-topics/hmh-efficacy/houghton_mifflin_harcourt_math_expressions_wi_case_studies_2013.pdf?la=en.

Appendix A continues from page 27-32.

Curriculum	Grade Level	Study
Math Expressions ***cont.***	Pre-K-2	45. Robin A. Honeycutt, "An Evaluation of an Elementary Mathematics Program at a Medium-Sized Suburban School District in North Carolina," dissertation, Wingate University, 2013, https://eric.ed.gov/?id=ED561826. 46. Janine T. Remillard, Barbara Harris, and Roberto Agodini, "The Influence of Curriculum Material Design on Opportunities for Student Learning," *ZDM Mathematics Education* 46, no. 5 (2014): 735-749 https://www.researchgate.net/publication/271660767_The_influence_of_curriculum_material_design_on_opportunities_for_student_learning. 47. "Impact Study: LaGrange School District 102, Illinois," HMH Professional Services, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HMH-Prof-Services-Impact-Study-LaGrange-IL.pdf. 48. "A Study of the Instructional Effectiveness of *Math Expressions*," report no. 42, Educational Research Institute of America, 2011, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HM-Math-Expressions_Intermediate-Efficacy_2011.pdf. 49. "A Study of the Instructional Effectiveness of Math Expressions Common Core 2013," report no. 473, Educational Research Institute of America, 2014, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HMH_MathExpressions_RM_2014.pdf. 50. Mark Dynarski et al., "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort," report to Congress, National Center for Education Evaluation and Regional Assistance, 2007.
Stepping Stones	Elementary (3-5)	51. Jordan Rickles and So Jung Park, *The Effect of Stepping Stones Math on Student Achievement*, American Institutes for Research, 2020, https://www.origoeducation.com/wp-content/uploads/2020/10/The-Effect-of-Stepping-Stones-on-Student-Achievement-AIR-Study-Report.pdf. 52. Forward Thinking Edu, *The Impacts of ORIGO Stepping Stones on Student Achievement: Magnolia Independent School District, Montgomery County, Texas*, report prepared for ORIGO Education, 2019, http://origoeducation-thailand.com/wp-content/uploads/2019/ORIGO-Education-MISD-White-Paper.pdf.
Utah middle school math	N/A	53. Ethan Prihar, Manaal Syed, Korinn Ostrow, Stacy Shaw, Adam Sales, and Neil Heffernan, "Exploring Common Trends in Online Educational Experiments," in proceedings of the 15th International Conference on Educational Data Mining (p. 27), 2022.
Various	N/A	54. Martyna Citkowicz, Jim Lindsay, David Miller, and Ryan Williams, "Heterogeneity in Mathematics Intervention Effects: Evidence from a Meta-Analysis of 191 Randomized Experiments," Society for Research on Educational Effectiveness, 2020.

Appendix A continues from page 27-32.

Curriculum	Grade Level	Study
Zearn	Combination	55. Jennifer Morrison, Betsy Wolf, Steven Ross, Kelsey Risman, and Caitlin McLemore, "Efficacy Study of Zearn Math in a Large Urban School District," Center for Research and Reform in Education, 2019. 56. Melissa Lambert and Joseph Sassone, "Accelerate, Don't Remediate: An Instructional Framework for Meeting the Needs of the Most Vulnerable Students after COVID School Closures," *Journal for Leadership and Instruction* 19, no. 2 (2020): 8–13. 57. Jennifer Knudsen, Patrik Lundh, Mindy Hsiao, and Daniela Saucedo, "Zearn Math Curriculum Study Professional Development Final Report," SRI International and TERC, 2020. 58. "Catching Up and Moving Forward," Zearn, 2022, https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf. 59. Alisa Szatrowski, "Technical Appendix: Efficacy Analysis of Zearn Math in Nebraska," Zearn, 2022, https://webassets.zearn.org/Implementation/NebraskaTechnicalAppendix.pdf. 60. Alisa Szatrowski, "Technical Appendix for: Catching Up and Moving Forward," Zearn, 2022, https://webassets.zearn.org/Implementation/accelerationmethodology.pdf. 61. Shirin A. Hashim, "Measuring the Effectiveness of Zearn Math in Louisiana," Zearn, 2021.

Appendix B. Summary of Study Outcomes

This summary of results from Bellwether's analysis shows an overwhelming number of studies that reported positive and mixed effects when assessing the impact of math curriculum on student success.

N=61

Outcomes	Number of Studies
Negative Effects	1
Nonsignificant Effects	4
Not Applicable	4
Mixed Effects	16
Positive Effects	36

Endnotes

1 U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2007, 2009, 2011, 2013, 2015, 2017, and 2019 Mathematics Assessments, retrieved February 17, 2021, from the Main NAEP Data Explorer, https://www.nationsreportcard.gov/ndecore/xplore/NDE.

2 Sarah Mervosh and Ashley Wu, "Math Scores Fell in Nearly Every State, and Reading Dipped on National Exam," *The New York Times*, October 24, 2022, https://www.nytimes.com/2022/10/24/us/math-reading-scores-pandemic.html.

3 Robert Balfanz, Liza Herzog, and Douglas J. Mac Iver, "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions," *Educational Psychologist* 42, no. 4 (2007): 223-235, https://new.every1graduates.org/wp-content/uploads/2012/03/preventing_student_disengagement.pdf.

4 Jaekyung Lee, "College for All: Gaps Between Desirable and Actual P-12 Math Achievement Trajectories for College Readiness," *Educational Researcher* 41, no. 2 (2012): 43-55, https://doi.org/10.3102/0013189X11432746.

5 Philip M. Sadler, Gerhard Sonnert, Zahra Hazari, and Robert Tai, "The Role of Advanced High School Coursework in Increasing STEM Career Interest," *Science Educator* 23, no. 1 (2014): 1-13, https://eric.ed.gov/?id=EJ1034751.

6 Lee, "College for All."

7 Ibid. See also Greg Duncan and Katherine Magnuson, "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems," 2011; Tyler Watts, Greg Duncan, Robert Siegler, and Pamela Davis-Kean, "What's Past Is Prologue," *Educational Researcher* 43, no. 7 (2014): 352-360, https://doi.org/10.3102/0013189X14553660.

8 U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2022 Mathematics National Student Group Scores and Score Gaps, retrieved November 18, 2022, https://www.nationsreportcard.gov/mathematics/nation/groups/?grade.

9 U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, "Data Point: Algebra I Coursetaking and Postsecondary Enrollment" (NCES 2019-154), https://nces.ed.gov/pubs2019/2019154.pdf.

10 Ibid.

11 U.S. Department of Education, "Algebra: Data Story," 2018, https://www2.ed.gov/datastory/stem/algebra/index.html.

12 Thomas J. Kane et al., *Teaching Higher: Educators' Perspectives on Common Core Implementation*, Center for Education Policy Research, February 2016, https://cepr.harvard.edu/files/cepr/files/teaching-higher-report.pdf.

13 Matthew R. Larson, "How to Recognize High-Quality Mathematics Content," Houghton Mifflin Harcourt, July 6, 2018, https://www.hmhco.com/blog/how-to-recognize-high-quality-mathematics-content.

14 Ibid.

15 Jere Confrey and Vicki Stohl, eds., *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations* (Washington, D.C.: The National Academies Press, 2004), https://doi.org/10.17226/11025.

16 Implementing a high-quality math curriculum is also a cost-effective approach to improving math outcomes for students. A study found that the average cost-effectiveness ratio for switching curriculum in math was 40 times greater, compared with reducing class size. Ulrich Boser, Matthew Chingos, and Chelsea Straus, *The Hidden Value of Curriculum Reform*, Center for American Progress, October 2015, https://cdn.americanprogress.org/wp-content/uploads/2015/10/06111518/CurriculumMatters-report.pdf.

17 "What Is the WWC? A Trusted Source About What Works in Education," Institute of Education Sciences, 2015, https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_info_what_061015.pdf.

18 Alan H. Schoenfeld, "What Doesn't Work: The Challenge and Failure of the What Works Clearinghouse to Conduct Meaningful Reviews of Studies of Mathematics Curricula," *Educational Researcher* 35, no. 2 (2006): 13–21, http://www.jstor.org/stable/3699951.

19 "How the EdReports Review Process Works," Review Tools, EdReports, https://www.edreports.org/process/review-tools.

20 Results from the articles reviewed yielded both positive and negative effects, along with nonsignificant and mixed effects. The summary of results shows an overwhelming number of studies that reported positive and mixed effects when assessing the impact of math curriculum on student success. Four of the 61 articles assessed the alignment and general features of the curriculum studied and were coded "Not Applicable" as they did not report on the impact of the curriculum studied. One study reported negative effects.

21 Matthew Swain, Bruce Randel, and Rebecca Norman Dvorak, "Impact Evaluation of Mathematics *i-Ready Instruction* for Elementary Grades Using 2018-19 Data: Final Report," no. 106. Human Resources Research Organization (HumRRO) for Curriculum Associates, 2020.

22 "ESSA Tiers of Evidence," REL Midwest, https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf.

23 Eduardo Hariton and Joseph J. Locascio, "Randomised Controlled Trials—the Gold Standard for Effectiveness Research," *BJOG: An International Journal of Obstetrics and Gynaecology* 125, no. 13 (2018): 1716, https://doi.org/10.1111/1471-0528.15199.

24 Ibid.

25 Richard J. Shavelson and Lisa Towne, eds., *Scientific Research in Education* (Washington, D.C.: National Academies Press, 2002), https://www.researchgate.net/profile/Richard-Shavelson/publication/284556232_Scientific_Research_in_Education/links/568d630d08aef987e565ee3e/Scientific-Research-in-Education.pdf.

26 Schoenfeld, "What Doesn't Work: The Challenge and Failure of the What Works Clearinghouse to Conduct Meaningful Reviews of Studies of Mathematics Curricula."

27 Ibid.

28 National Academies of Sciences, Engineering, and Medicine, 2001, *Adding It Up: Helping Children Learn Mathematics*, Washington, DC: The National Academies Press, https://doi.org/10.17226/9822.

29 Schoenfeld, "What Doesn't Work: The Challenge and Failure of the What Works Clearinghouse to Conduct Meaningful Reviews of Studies of Mathematics Curricula."

30 James Hiebert and Douglas A. Grouws, "The Effects of Classroom Mathematics Teaching on Students' Learning," in Frank K. Lester Jr., ed., *Second Handbook of Research on Mathematics Teaching and Learning* (Charlotte, NC: Information Age, 2007): 371–404.

31 Leib Sutcher, Linda Darling-Hammond, and Desiree Carver-Thomas, "Understanding Teacher Shortages: An Analysis of Teacher Supply and Demand in the United States," *Education Policy Analysis Archives* 27, no. 35 (2019), https://doi.org/10.14507/epaa.27.3696.

32 Thomas J. Kane, "Never Judge a Book by Its Cover—Use Student Achievement Instead," Brookings, March 3, 2016, https://www.brookings.edu/research/never-judge-a-book-by-its-cover-use-student-achievement-instead/.

33 Roberto Agodini, Barbara Harris, Neil Seftor, Janine Remillard, and Melissa Thomas, "After Two Years, Three Elementary Math Curricula Outperform a Fourth" (ED544185), NCEE evaluation brief, 2013, https://files.eric.ed.gov/fulltext/ED544185.pdf.

34 Ibid.

35 Jennifer Knudsen, Patrik Lundh, Mindy Hsiao, and Daniela Saucedo, *Zearn Math Curriculum Study Professional Development Final Report*, SRI Education and TERC, December 2020, http://webassets.zearn.org/Implementation/ZearnMath_CurriculumStudyPD_FinalReport.pdf.

36 Confrey and Stohl, *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations*.

37 *Curriculum Implementation Guide*, Chiefs for Change, November 2019, https://chiefsforchange.org/wp-content/uploads/2020/07/Curriculum-Implementation-Guide.pdf.

38 JEM & R LLC, "HMH Into Math: K-8 Early Outcomes 2020SY," Houghton Mifflin Harcourt, 2020, https://s3.amazonaws.com/prod-hmhco-vmg-craftcms-public/research/HMH-Into-Math-Research-Results-Paper-2019-2020-SY.pdf.

39 Alisa Szatrowski, "Technical Appendix: Efficacy Analysis of Zearn Math in Nebraska," Zearn, 2022, https://webassets.zearn.org/Implementation/NebraskaTechnicalAppendix.pdf.

40 Russell Gersten, Mary Jo Taylor, Tran D. Keys, Eric Rolfhus, and Rebecca Newman-Gonchar, *Summary of Research on the Effectiveness of Math Professional Development Approaches*, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast, January 2014, https://files.eric.ed.gov/fulltext/ED544681.pdf.

41 "Internationally, U.S. Stands in Middle of Pack on Science, Math, Reading Scores," Pew Research Center, February 11, 2019, https://www.pewresearch.org/fact-tank/2017/02/15/u-s-students-internationally-math-science/ft_17-02-14_stem_dot-2/.

42 Jill Denner, Olivia Valdes, Daniel J. Dickson, and Brett Laursen, "Math Interest and Self-Concept Among Latino/a Students: Reciprocal Influences Across the Transition to Middle School," *Journal of Adolescence* 75, no. 1 (2019): 22-36, https://doi.org/10.1016/j.adolescence.2019.06.015.

43 Jo Boaler, Jack A. Dieckmann, Graciela Pérez-Núñez, Kathy Liu Sun, and Cathy Williams, "Changing Students Minds and Achievement in Mathematics: The Impact of a Free Online Student Course," *Frontiers in Education* 3 (2018), https://www.frontiersin.org/articles/10.3389/feduc.2018.00026/full.

44 *Envision Math 2.0 Research Overview: Evidence of Effectiveness*: A Summary of the 2015-2017 Longitudinal Efficacy Study, Savvas Learning Company, 2020, https://assets.savvas.com/asset_mgr/current/202038/enVm2_RCT_Research_Overview_Savvas.pdf.

45 Jordan Rickles and So Jung Park, *The Effect of Stepping Stones Math on Student Achievement*, American Institutes for Research, 2020, https://www.origoeducation.com/wp-content/uploads/2020/10/The-Effect-of-Stepping-Stones-on-Student-Achievement-AIR-Study-Report.pdf.

46 M. Cameron Hay, ed., M*ethods That Matter: Integrating Mixed Methods for More Effective Social Science Research* (Chicago, IL: University of Chicago Press, 2015).

47 Katie Pak, Morgan S. Polikoff, Laura M. Desimone, and Erica Saldívar García, "The Adaptive Challenges of Curriculum Implementation: Insights for Educational Leaders Driving Standards-Based Reform," *AERA Open* 6, no. 2 (2020).

48 Gülçin Karakus, "A Literary Review on Curriculum Implementation Problems," *Shanlax International Journal of Education* 9, no. 3 (2021): 201-220.

49 Cynthia E. Coburn, William R. Penuel, and Caitlin C. Farrell, "Fostering Educational Improvement with Research-Practice Partnerships," *Phi Delta Kappan* 102, no. 7 (2021): 14-19, https://journals.sagepub.com/doi/full/10.1177/00317217211007332.

About the Authors

THOMAS GOLD

Thomas Gold is a senior associate partner at Bellwether in the Policy and Evaluation practice area. He can be reached at **thomas.gold@bellwether.org**.

KRISTEN CARROLL

Kristen Carroll is an associate partner at Bellwether in the Policy and Evaluation practice area. She can be reached at **kristen.carroll@bellwether.org**.

LEONARD D.T. NEWBY

Leonard D.T. Newby is a senior analyst at Bellwether in the Policy and Evaluation practice area. He can be reached at **leonard.newby@bellwether.org**.

MELISSA STEEL KING

Melissa Steel King is a partner and Evaluation leader in the Policy and Evaluation practice area at Bellwether. She can be reached at **melissa.king@bellwether.org**.

About Bellwether

Bellwether is a national nonprofit that exists to transform education to ensure systemically marginalized young people achieve outcomes that lead to fulfilling lives and flourishing communities. Founded in 2010, we work hand in hand with education leaders and organizations to accelerate their impact, inform and influence policy and program design, and share what we learn along the way. For more, visit **bellwether.org**.

ACKNOWLEDGMENTS

We would like to thank the many individuals who gave their time and shared their knowledge with us to inform our work, including the Bill & Melinda Gates Foundation for its financial support of this project. We want to thank Paul Cobb, Barbara Harris, Melanie Manley, John F. Pane, Janine Remillard, Toni Smith, and Trena Wilkerson for being generous with their time and participating in interviews, and Barbara Harris for providing feedback on an initial draft of this analysis.

We would also like to thank Bellwether colleague Cathy Thomas for her support. Thank you to Alyssa Schwenk, Andy Jacob, Zoe Campbell, Julie Nguyen, and Amber Walker for shepherding and disseminating this work, and to Super Copy Editors.

The contributions of these individuals and entities significantly enhanced our work; however, any errors in fact or analysis remain the responsibility of the authors.