



Impact of the Illustrative Math Curriculum on Math Achievement

A Matched-Comparison
Study in Massachusetts



Noman Khanani, Dalia Zabala, Shannon
Stackhouse, Jena Gu, and Kirk Walters

July 2025

© 2025 WestEd. All rights reserved.



Suggested citation: Khanani, N., Zabala, D., Stackhouse, S., Gu, J., & Walters, K. (2025). *Impact of the Illustrative Math Curriculum on math achievement*. WestEd.

WestEd is a nonpartisan, nonprofit organization that aims to improve the lives of children and adults at all ages of learning and development. We do this by addressing challenges in education and human development, increasing opportunity, and helping build communities where all can thrive. WestEd staff conduct and apply research, provide technical assistance, and support professional learning. We work with early learning educators, classroom teachers, local and state leaders, and policymakers at all levels of government.

For more information, visit [WestEd.org](https://www.wested.org). For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the E-Bulletin, our semimonthly e-newsletter, at [WestEd.org/subscribe](https://www.wested.org/subscribe).

This report was made possible through support from the Gates Foundation.

Contents	
Executive Summary	1
Introduction	3
Massachusetts Context	4
What the Study Examined	5
Data Sources	5
Methodology	7
Research Question 1: School Characteristics	7
Research Question 2: Fidelity of Implementation	7
Research Question 3: Impact of IM Adoption	7
Study Findings	8
Research Question 1: School Characteristics	8
Research Question 2: Implementation Experience	9
Research Question 3: Impact of IM on Student Math Achievement	10
Implications of the Findings	16
Limitations of the Findings	16
Future Directions for Related Research	17
References	18
Appendix: Additional Methodology Details and Supplemental Findings	19

Executive Summary

This study is the first to explore the impact of the Illustrative Math (IM) curriculum on student math performance. IM is one of the nation's most widely adopted high-quality math curricula.



Using survey, state, and public town data from districts in Massachusetts, researchers explored how student math performance changed in 71 schools after they began implementing the IM curriculum in middle school grades compared to similar schools using a different high-quality math curriculum.

- On average, IM schools did not perform significantly differently from their counterparts over time after IM implementation.
- However, schools that adopted IM before the COVID-19 pandemic performed significantly better than their counterparts, particularly in grades 7 and 8, by the 3rd year of implementation.
- While the comparison schools experienced significant learning loss following the onset of the COVID-19 pandemic, the IM schools experienced no decline in math achievement during this time.

- In contrast, schools that adopted IM during or after the pandemic showed no meaningful improvement relative to similar schools that implemented other high-quality math curricula.

The difference between schools that adopted IM before the pandemic and those that adopted IM after it likely reflects implementation challenges during COVID-19, including disrupted professional development, remote learning complications, and reduced collaborative planning time, which are all critical for successful curriculum adoption.

The study focused primarily on higher performing, less economically disadvantaged schools, so results may not apply to all contexts. Still, the findings suggest that with proper implementation support and stable conditions, IM can deliver meaningful academic gains.

For education leaders considering curriculum adoption, this research underscores that *how* and *when* you implement matters as much as *what* you implement. High-quality curricula require sustained professional development and supportive implementation conditions to reach their potential.

Introduction

Nationally, the high-quality instructional materials (HQIM) movement has gained momentum over the past decade, driven by research showing that access to rigorous standards-aligned curricula can significantly impact student achievement (Kane et al., 2016; Polikoff, 2021; Sahm, 2015). Several states have established initiatives identifying high-quality programs (often using reviews from EdReports or their own review panels), incentivizing their adoption, and investing in professional learning to support implementation (Doan et al., 2022). These efforts are often supported by philanthropic organizations and coordinated through networks like the Council of Chief State School Officers and Instruction Partners.

Illustrative Math (IM) has emerged as one of the most widely adopted HQIM programs in math across the United States, particularly in grades 6–8 and increasingly in K–5 (Diliberti et al., 2023). Known for its strong alignment with rigorous college- and career-readiness standards, IM emphasizes conceptual understanding, coherence, and student-centered problem-solving. It consistently receives top marks from both EdReports and state-level reviews.

Despite this wide adoption, there remains no causal evidence to date on the impact of IM implementation on student outcomes.¹ Using data from Massachusetts, this study explores how math achievement changed in schools following IM adoption compared to a matched set of schools using different curricula. Additionally, the study documents the characteristics of schools that use IM and district-reported perceptions of implementation.

¹ A matched comparison study of Amplify Desmos Math, which significantly incorporated IM lesson materials, found positive effects on student math achievement (McKinney et al., 2023).

Massachusetts Context

In 2018, the Massachusetts Department of Elementary and Secondary Education (DESE) began an HQIM initiative as part of the state’s broader effort to promote equity and academic excellence by ensuring all students have access to rigorous, standards-aligned curricula. The initiative supports districts in selecting, adopting, and implementing materials grounded in research and designed to meet the diverse needs of all learners. A key component of this work is the CURATE (CURriculum RATings by TEachers) project, launched to provide DESE-specific evaluations of commonly used instructional materials, complementing national reviews like EdReports.²

CURATE panels, made up of Massachusetts educators, assess curricula in subjects like math and English based on content alignment, support for all learners, and usability. These reviews help guide district purchasing decisions and inform DESE’s professional development (PD) and implementation supports, positioning HQIM and CURATE as central pillars in the state’s strategy for accelerating student learning and improving instructional quality. IM—as provided by McGraw Hill, Imagine Learning, Open Up Resources, and Kendall Hunt—has met all expectations in the CURATE rubrics. The state’s most recent curriculum report in 2024, which documents the core curriculum self-reported by districts in each subject, demonstrates that 20 percent of districts list IM as their core math curriculum in at least one middle school grade.³ An additional 20 percent simply note “Multiple Materials,” which may include IM.

² EdReports evaluates core instructional materials using educator-developed rubrics that assess alignment with college- and career-ready standards, usability, and other key quality indicators. These reviews are conducted by teams of experienced educators who analyze materials over several months, gathering evidence and reaching consensus on scoring recommendations. The organization’s reports are publicly available and widely used by school districts and states to inform curriculum adoption decisions.

³ School grade spans vary in the state. The sample of IM schools in this study included a mix of K–5/6, 6–8, and 6/7–12 schools. Though the same curriculum was often used by all grades within a school, there were instances in which the curriculum varied by grade.

What the Study Examined

This study addressed the following research questions:

1. What are the characteristics of schools in Massachusetts where IM is implemented in a middle school grade compared to schools that implement other curricula?
2. To what extent is IM implemented with fidelity, according to district leaders?
3. What is the impact of IM adoption on student math achievement? Does the impact vary over time, by year of adoption, or by grade level?

Data Sources

For Research Question 1, the research team identified schools that listed IM as their middle school math core curriculum based on DESE's most recent public report on the math curricula used by schools at each grade level throughout the state. State administrative data on school demographic characteristics and math

performance during the 2023/24 school year were used to compare IM districts to non-IM districts.

The research team excluded schools that listed "Multiple Materials" and "Open Up Resources" as their middle school math curriculum because it could not be discerned whether IM was among the curricula used. Excluded schools were more likely to be lower performing, be based in urban contexts, and have a higher percentage of students from households with low income than were included schools. A total of 101 schools across 55 districts were recorded as using IM during the 2024/25 school year.



For Research Questions 2 and 3, the research team sent a short survey to math leaders in districts where IM was listed as implemented in at least one middle school. The survey included basic questions about IM use, such as the version and year of IM adoption. Additionally, district leaders were asked to provide an estimate of the proportion of time in which IM was used in a typical week by teachers. The survey also provided space for

district leaders to reflect on IM use more generally. Initial outreach began in March 2025 and was conducted via email to identified math or curriculum directors. Follow-up occurred via email and phone calls. The survey closed in May 2025 with 50 percent of district leaders contacted having responded to the survey ($n = 28$).

For districts from which no response to the survey was received, to address Research Question 3, the research team identified the IM start year in district middle schools by inspecting district and school websites along with public town meeting minutes and reports. Overall, the research team was able to identify the IM start year for 85 of the 101 schools serving a middle school grade in Massachusetts where IM was specifically listed as the core math curriculum. Table 1 below shows the number of schools and districts that adopted IM by year.

Table 1. IM Adoption in Massachusetts Over Time

Year	2018/19	2019/20	2020/21	2021/22	2022/23	2023/24	2024/25
Number of schools	19	5	1	14	22	10	14
Number of districts	4	3	1	10	13	9	4

The research team used this information to generate a grade-by-school panel data set spanning the 2016/17 to 2023/24 school years. The data set included school demographic information, math performance on the state standardized assessment (the Massachusetts Comprehensive Assessment System, or MCAS), the math curriculum currently used, and the IM start year. Outcomes could not be analyzed for the 2024/25 cohort of schools because 2025 spring test data were not available at the time of the analyses; this resulted in a final IM sample of 71 schools. Additionally, for Research Question 3 analyses, the 1st year of IM implementation for the 2019/20 cohort was recoded to 2020/21 because state assessments were not administered in spring 2020.

Methodology

Research Question 1: School Characteristics

For Research Question 1, the research team compared IM districts to non-IM districts on spring 2024 MCAS math scaled scores and on the percentage of students identified as low income, students with disabilities, English Learners, and students within each racial category. Two groups were selected as comparison groups: those implementing another HQIM as defined by DESE and those using a non-HQIM curriculum. We noted both statistical significance and effect size differences between the groups.

Research Question 2: Fidelity of Implementation

For Research Question 2, we provided simple frequencies of the select-response survey questions. Additionally, we summarized notable themes that were mentioned in the open response questions and included direct quotes for more context.

Research Question 3: Impact of IM Adoption

To examine IM's impact on student math achievement, we compared each grade within IM schools to a matched group of schools using other high-quality math programs. We matched schools on prior achievement and key demographics to make the groups as similar as possible before IM was introduced. We then compared their math scores in the first 3 years after IM implementation, adjusting for baseline differences. Finally, we explored whether IM's impact varied by grade level and whether the program started before or after the pandemic. More information on the methodology is provided in the Appendix.

Study Findings

Main findings from the analyses are summarized below. Supplemental findings can be found in the Appendix.

Research Question 1: School Characteristics

To understand which types of schools implement IM in the state, we compared the demographic and student performance characteristics of IM, other HQIM, and non-HQIM schools.

IM schools statistically differed from other HQIM schools and non-HQIM schools on several characteristics.

Table 2 summarizes the demographic and student performance differences between IM schools, schools implementing other HQIM, and non-HQIM schools.

Table 2. Differences Between IM and Non-IM schools, 2023/24

Characteristic	IM schools	HQIM, not IM schools	Non-HQIM schools
Percentage of students achieving proficiency on math MCAS	39%	34%**	43%**
Math z-score ^a	0.15	-0.09***	0.32**
English Learners	12%	12%	9%***
Students from households with low income	40%	48%***	38%
Students with disabilities	20%	20%	20%

Characteristic	IM schools	HQIM, not IM schools	Non-HQIM schools
Black students	9%	15%***	8%
White students	53%	53%	59%**
Hispanic students	23%	24%	21%
Total grade-school observations	250	294	531
Total schools	101	142	238

Note. Asterisks indicate the statistical significance of the value in the cell compared to the corresponding row value in the “IM Schools” column. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.

^a The z-scores of math scale scores show how a school’s average math score compares to the overall state average. A z-score of 0 means the school is exactly at the average, positive scores are above average, and negative scores are below average.

Compared to other HQIM schools, IM schools were higher performing on the state math assessment and had a lower percentage of Black students and students from low-income backgrounds. However, relative to non-HQIM schools (excluding schools where the curriculum used was unknown), IM schools were slightly lower performing and more racially diverse.

Research Question 2: Implementation Experience

This research question asked district leaders to indicate their experience with IM implementation and how frequently IM was used in the core classroom.

IM was most frequently used in grade 6.

Of the 28 district representatives that responded to the survey,⁴ all but one indicated that IM was used in grade 6, and 71 percent reported using IM in grades 7 and 8. Kendall Hunt was the most frequently cited version (43%), followed by Imagine Learning (32%), McGraw Hill (18%), and Open Up (7%).

⁴ Two districts noted that they no longer implemented IM. These districts were excluded from the sample because we could not discern when they stopped using IM.

Districts reported high usage rates of IM.

During the 2024/25 school year, most districts (75%) noted that IM was being used 76–100 percent of the time, on average, in classrooms; 21 percent reported it was being used 51–75 percent of the time; and only one said it was being used between 26 and 50 percent of the time. Most respondents indicated that these patterns of usage applied to prior years.

Professional development is important for supporting implementation.

The 1st year of implementation was mentioned by multiple districts as a pilot year. In these districts, additional PD was established in the 2nd year to improve fidelity. In two cases, districts shifted from one version to another between school years, citing costs as an important factor. Several respondents noted the importance of PD in improving implementation, though districts had varying experiences with PD to date.

Districts experienced common challenges with implementation.

Multiple district respondents shared challenges that suggest IM effectiveness could vary by student proficiency levels. For example, one respondent said teachers struggle to catch students up when they are absent. Another respondent felt that it was difficult to differentiate IM for the range of student proficiency in classrooms. Finally, one respondent said that teachers have expressed concern about the limited set of practice problems and traditional quizzes in the curriculum. These comments imply that IM may not be equally effective across varying student populations.

Research Question 3: Impact of IM on Student Math Achievement

The following findings summarize the differences between IM schools and their matched HQIM counterparts on math achievement over time:

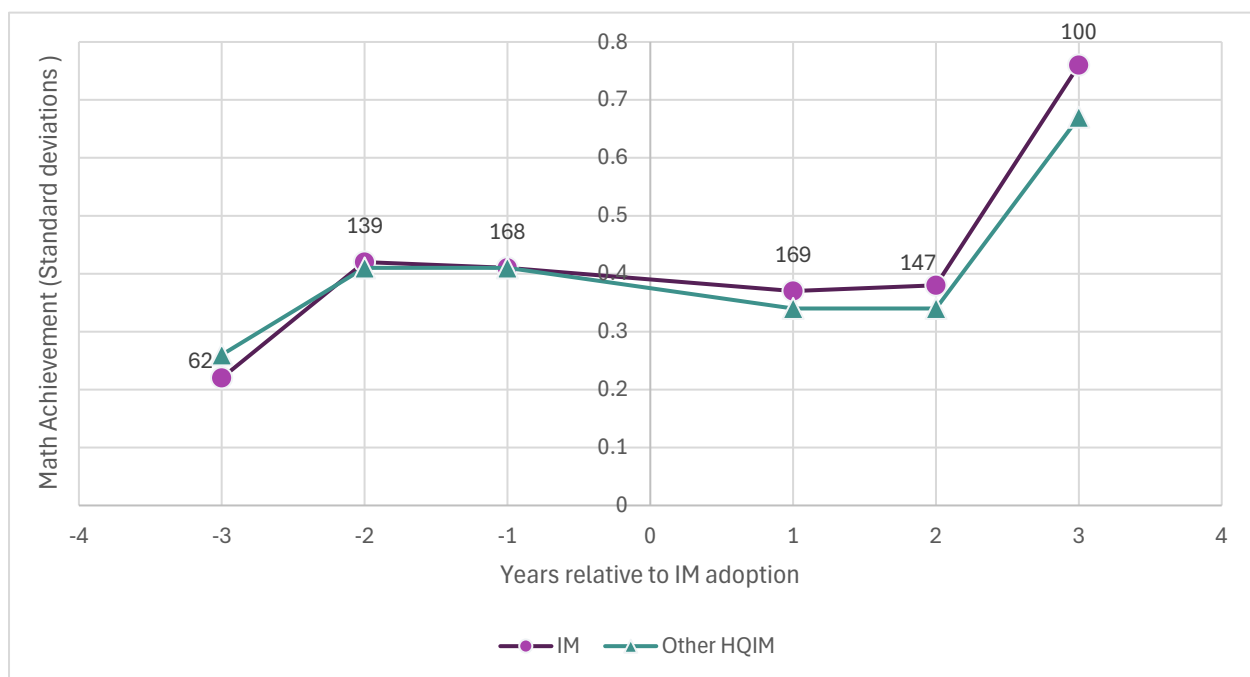
- Schools that adopted IM in Massachusetts between 2020 and 2024 performed, on average, no better on middle school math performance compared to schools that used other HQIM curricula.
- Schools that adopted IM before the COVID-19 pandemic began, however, performed significantly better in grade 8 math during the 1st and 3rd years of

adoption. By the 3rd year of adoption, these schools were also doing significantly better on grade 7 math performance.

There was no impact of IM adoption on student math achievement until 3 years after implementation.

Figure 1 presents the math achievement trend before and after IM implementation for IM schools versus their matched group of other HQIM schools. In the 3 years prior to IM implementation, IM schools and their matched HQIM counterparts were performing at similar levels in math. By the 3rd year following IM Implementation, however, the average math achievement in IM schools was slightly better than their HQIM counterparts. Adjusting for prior characteristics, the difference was negligible in the first 2 years, but by the 3rd year, the effect grew to 0.19 standard deviations and was statistically significant (Table A2).

Figure 1. Schools that adopted IM performed slightly better than their matched HQIM schools, particularly by the 3rd year of implementation.



Note. The y-axis indicates math scores operationalized as z-scores. The x-axis represents the year centered around the start of IM implementation such that -1 is the spring score before IM implementation and 1 is the spring score for the 1st year of IM implementation. For example, for schools that began IM implementation in 2021/22, -1 would correspond to spring 2021, and 1 would correspond to spring 2022. Data labels indicate the *n* size of the treatment group at each time point.

Though the figure suggests that both IM and the matched HQIM schools experienced large increases in achievement by the 3rd year of IM implementation, this is simply due to which observations are included in the 3rd-year estimate. The 2022/23 and 2023/24 cohorts—which make up 40 percent of the sample—are excluded from the 3rd-year estimate because, at the time of analysis, they had not yet been assessed 3 years after implementation (see data labels in Figure 1 for n sizes). Schools from these two cohorts were lower performing than the earlier cohorts. **Overall, the trend for both IM and HQIM schools between the immediate year before implementation and each year following is stable or shows a slight decrease, likely attributable to learning loss associated with the COVID-19 pandemic.**

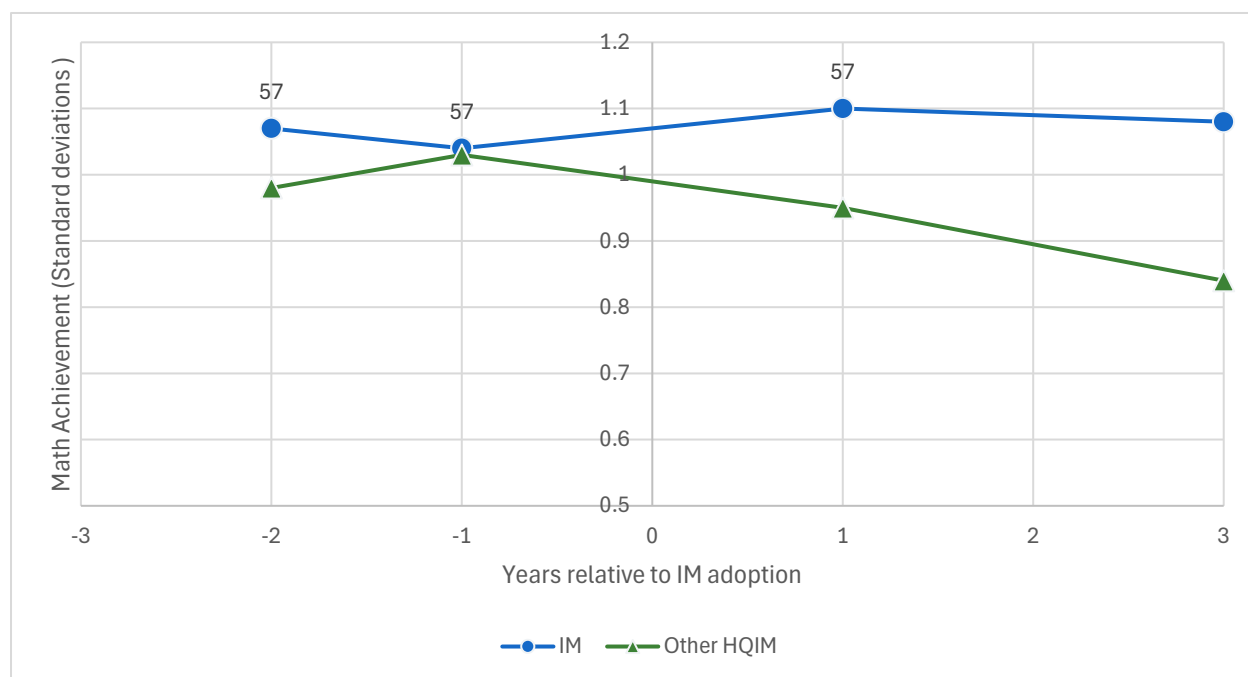
Impact of IM on Math Achievement, by Cohort

Given potential pandemic-related disruptions to PD and implementation, the impact of IM may have varied depending on whether schools adopted IM before or after COVID-19. For instance, PD after spring 2020 was more likely to be virtual, which could have impacted IM implementation.

Schools that adopted IM before the COVID-19 pandemic showed relative gains in math by maintaining performance as their matched schools declined.

Figure 2 displays the trends for the 19 schools that adopted IM in 2018/19, the year before COVID-19, compared to their matched HQIM counterparts. Notably, as indicated by the two pre-IM math score measures (spring 2016 and spring 2017), the pre-COVID-19 IM cohort included high-achieving schools. Yet we were still able to retrieve a similarly high-achieving matched comparison group.

Figure 2. Schools that adopted IM before COVID-19 did not experience learning loss, but achievement in comparison schools decreased significantly.



Note. The y-axis indicates math scores operationalized as z-scores. The x-axis represents the year centered around the start of IM implementation such that -1 is the spring score before IM implementation and 1 is the spring score for the 1st year of IM implementation. Because the pre-COVID-19 cohort all started IM at the same time, 1 corresponds to 2018/19 and -1 corresponds to 2017/18. Because spring 2020 scores were not available, no IM Year 2 outcomes are shown. Data labels indicate the n size of the treatment group at each time point.

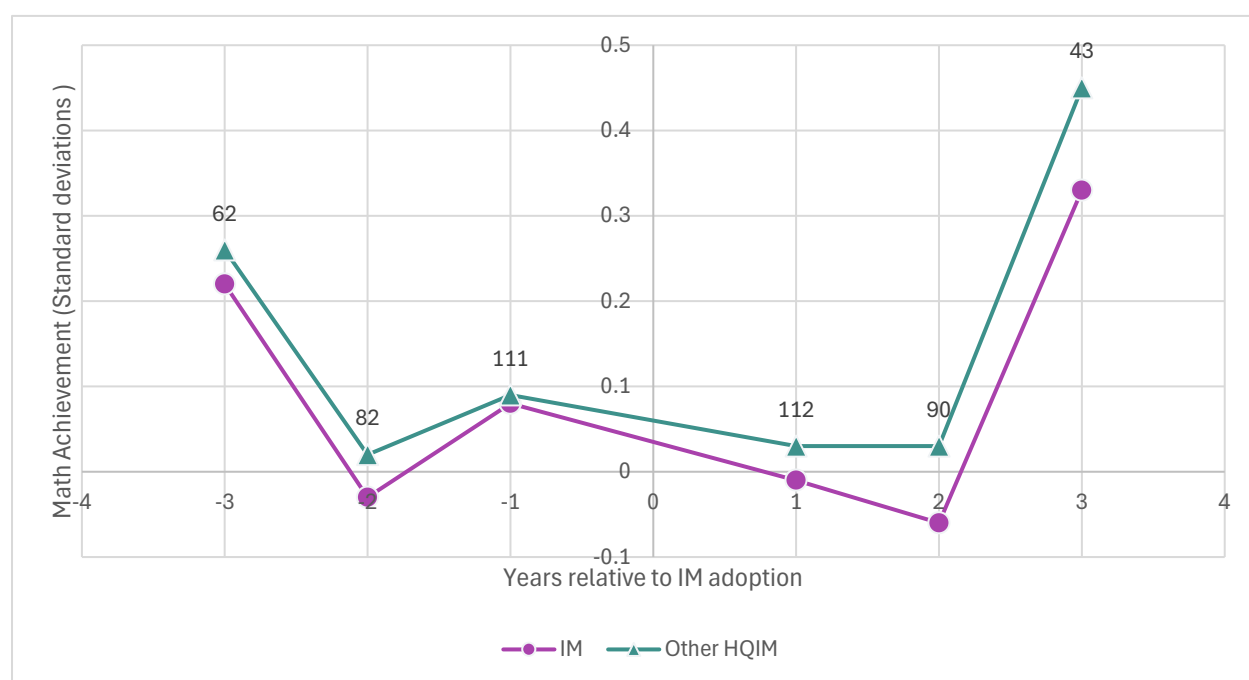
Adjusting for prior characteristics, the IM impact on student math achievement for the pre-COVID-19 cohort was 0.18 standard deviations in the 1st year and 0.39 standard deviations in the 3rd year (Table A2). Both impacts were statistically significant.

These impacts are not, however, a result of significant gains made by the IM schools. The IM schools experienced largely stagnant achievement during the post-IM period, but scores significantly declined for the matched comparison group. Though some of this decline came after COVID-19, scores in the other HQIM schools also decreased between spring 2018 and spring 2019, when the IM schools were in their 1st year of implementation. This finding is consistent with the impact results in McKinney et al. (2023) and other studies documenting learning related to COVID-19 disruptions (e.g., Kuhfeld et al., 2022).

Schools that adopted IM after the start of the COVID-19 pandemic performed no differently from their matched HQIM schools, indicating no measurable impact of IM.

Figure 3 below presents the same before-and-after IM trend in math achievement for the remaining group of 52 schools that adopted IM after the start of the pandemic. The differences in post-IM scores between the IM and matched other HQIM group suggest a slight negative effect, though not statistically significant. The adjusted effect sizes are -0.04, -0.11, and -0.06 standard deviations for Years 1, 2, and 3, respectively (Table A2).

Figure 3. Schools that adopted IM after COVID-19 did not perform significantly differently following implementation relative to their matched comparison schools.



Note. The y-axis indicates math scores operationalized as z-scores. The x-axis represents the year centered around the start of IM implementation such that -1 is the spring score before IM implementation and 1 is the spring score for the 1st year of IM implementation. For example, for schools that began IM implementation in 2021/22, -1 would correspond to spring 2021, and 1 would correspond to spring 2022. Data labels indicate the *n* size of the treatment group at each time point.

Like the pooled results in Figure 1, the large increase in Year 3 scores is due to the omission of the last two cohorts because their scores were not observed that far out at the time of analyses. **The main trend suggests a decline in scores in both groups following IM implementation, which largely corresponds with postpandemic trends.** On average, scores largely remained stagnant following the initial decline, whereas there is

some evidence of more recovery in the matched comparison group; but this difference is not statistically significant.

Impact by Grade Level

Finally, we explored the impact of IM on math achievement by grade level overall and within cohort. The adjusted results, shown in Table 3, indicate important differences in impact by grade level.

The overall effect of IM was largely driven by grades 7 and 8 gains in the pre-COVID cohort.

For the pre-COVID-19 cohort, a large and statistically significant effect of 0.40 standard deviations was apparent by the 1st year of IM adoption in grade 8. This grade 8 effect was still present in the 3rd year, along with a large and statistically significant effect of 0.50 standard deviations in grade 7. The impact in grade 6 was modest at 0.16 and 0.18 standard deviations in Years 1 and 3, respectively, but not statistically significant.

Table 3. IM Impact in Math Achievement Over Time, by Grade and Cohort

Year	Grade	All IM schools	Pre-COVID-19 cohort	Post-COVID-19 cohort
Year 1	6	0.05	0.16	-0.01
Year 1	7	-0.02	-0.01	-0.03
Year 1	8	0.07	0.40***	-0.07
Year 2	6	-	-	-0.20*
Year 2	7	-	-	0.03
Year 2	8	-	-	-0.14
Year 3	6	0.04	0.18	-0.10
Year 3	7	0.30***	0.50***	0.00

Year	Grade	All IM schools	Pre-COVID-19 cohort	Post-COVID-19 cohort
Year 3	8	0.20*	0.43***	-0.10

Note. Cell values indicate the adjusted difference between IM and matched HQIM observations in standard deviation units. The “-” indicates fields for which estimates were not available because Year 2 estimates were only available for the post-COVID-19 cohort because spring 2020 corresponded to Year 2 for the pre-COVID-19 cohort, when assessments were not administered. Asterisks indicate statistical significance: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.

The impact estimates for the post-COVID-19 cohort reveal a similar effect as the aggregate effects discussed above, with almost all effects being negligible and not statistically significant. The one exception is a statistically significant negative effect ($p < 0.10$) of 0.20 standard deviations in grade 6 during the 2nd year of adoption, though this effect decreased to -0.10 by the 3rd year and was no longer statistically significant.

Implications of the Findings

The study’s findings suggest that implementation timing may have played a critical role in the effectiveness of IM. As implied by some of the survey responses, districts that adopted IM in the early years of the pandemic faced significant disruptions to PD, coaching, collaborative planning, and classroom routines that are all typically necessary for high-fidelity implementation. Remote learning, staff turnover, and shifting instructional priorities may have limited the opportunities for teachers to fully engage with the curriculum and adapt their practice.

Now that school operations have returned to more stable footing, districts may find greater success with IM by reinvesting in the foundational supports—such as sustained PD, instructional leadership, and collaborative time—that are essential for effective curriculum use. The strong outcomes observed in the earliest implementation cohort suggest that, with adequate support, IM can still produce meaningful gains, particularly in upper middle grades.

Limitations of the Findings

There are multiple important limitations to consider when interpreting these findings. Although the study used a matched comparison design to account for observable differences between IM-adopting schools and others, unmeasured factors related to both IM adoption and math achievement could still explain some of the observed effects. While a

randomized controlled trial would have provided the most confident impacts of IM implementation on student achievement, such an undertaking was not possible for this study. Still, the study was designed and executed to meet *What Works Clearinghouse (WWC) Standards 5.1 With Reservations*.

The analysis also had limited statistical power because the data were aggregated at the grade–school level rather than the student level, reducing sensitivity to detect smaller effects. Based on WWC standards, an effect of 0.25 standard deviations is considered practically significant, and, as such, this benchmark was used to highlight important differences.

Additionally, there is uncertainty about how well IM was implemented. Though most districts surveyed reported high usage of IM, this measure of fidelity was captured through a single self-reported item in a survey completed by half of the districts reported by DESE as using IM. Furthermore, this measure did not capture the quality of implementation. IM’s leadership team has hypothesized that high-quality implementation is critical to improving student math achievement and has created a detailed implementation reflection tool to measure such quality. Additionally, responses were based on broad estimates from district personnel who may have varying knowledge of classroom practices.

Finally, the sample primarily included schools in small suburban districts with students who are higher achieving and less economically disadvantaged. Many urban schools with higher concentrations of poverty were excluded because they used multiple math curricula, making it difficult to isolate the impact of IM. As a result, the findings may not generalize to more diverse or underresourced school contexts.

Future Directions for Related Research

To provide a more comprehensive and confident estimate of IM impacts on student achievement, future studies could focus on sites with more socioeconomic and racial diversity. Replication studies in other states can also help inform whether these findings hold true in other contexts outside of Massachusetts. Additionally, reexamining impacts in Massachusetts for more recent and future cohorts of IM adopters can test this study’s proposed theory of IM impacts being potentially stronger in contexts where the pandemic impact is less consequential. Finally, including schools with well-documented measures of IM implementation fidelity and quality would strengthen the findings.

References

- Diliberti, M., Destler, K., Rainey, L., & Schwartz, H. (2023). How are district leaders thinking about mathematics? *Selected findings from the American School District panel*. RAND Corporation.
- Doan, S., Kaufman, J. H., Woo, A. S. H. L. E. Y., Tuma, A. P., Diliberti, M. K., & Lee, S. A. B. R. I. N. A. (2022). *How states are creating conditions for use of high-quality instructional materials in K–12 classrooms*.
- Kane, T., Owens, A., Marinell, W., Thal, D., & Staiger, D. (2016). *Teaching higher: Educators' perspectives on common core implementation*. Center for Education Policy Research.
- Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, 51(7), 500–506.
- McKinney, D., Strother, S., Walters, K., & Schneider, S. (2023). *The effect of Desmos Math Curriculum on middle school mathematics achievement in nine states*. WestEd.
- Polikoff, M. (2021). *Beyond standards: The fragmentation of education governance and the promise of curriculum reform* (Vol. 8). Harvard Education Press.
- Sahm, C. (2015). Curriculum counts: NYC public schools and the Common Core. *Civic Report. Manhattan Institute*, 99.

Appendix: Additional Methodology Details and Supplemental Findings

Research Question 3 was designed to meet *What Works Clearinghouse (WWC) Standards (v.5.1) With Reservations* concerning the impact of IM implementation on student math achievement, operationalized as a z-score.⁵ Since IM schools may differ from non-IM schools at baseline (i.e., before IM implementation) on demographic characteristics and achievement measures, we constructed a matched comparison group. The level of observation is at the grade–school level, which refers to the average performance of a grade level within a particular school as opposed to the overall performance of the school. Each observation in the data set represented aggregate performance at the grade–school level instead of the overall school level because IM may not have been implemented at each grade level or the start years differed by grade level. By including each grade within a school as a separate observation, we can account for differences in start years and implementation status.

We matched each grade–school observation from an IM school to one HQIM and one non-HQIM grade–school observation using propensity score matching. This methodology matches IM observations to non-IM observations based on their probability (i.e., propensity) to implement IM based on selected variables. As per WWC standards, we selected one prior achievement measure (math scaled scores on the state spring assessment) and two demographic characteristics (percentage of students identified as low income and percentage of students with disabilities) with the aim of reducing differences between IM and non-IM observations on these variables to less than 0.25 standard deviations. Though more variables are typically recommended for propensity score matching, the limited

⁵ The z-scores here represent transformations of the math scaled scores to have a mean of 0 and a standard deviation of 1 within grade and year.

sample size and stark differences between IM and non-IM schools made it difficult to identify appropriate matches when several covariates were used.

Since schools implemented IM in stages between 2018/19 and 2023/24, we conducted matching separately by both grade and cohort. This ensured that each observation was exactly matched at the grade level and that the corresponding prior achievement measure reflected the precise year before IM implementation. For example, the grade–school observations that adopted IM in 2018/19 were matched to a set of non-IM grade–school observations based on the values of the variables during the 2017/18 school year. Whereas for the 2022/23 IM cohort, the matches are based on values of the variables for the 2021/22 school year.

Table A1 provides evidence of baseline equivalence between the IM and matched other HQIM observations. Differences between IM and the matched group of other HQIM observations on math achievement, proportion of students from households with low income, and proportion of students with disabilities were all below 0.25 standard deviations.

Table A1. Baseline Equivalence Between IM and Matched HQIM Observations

Characteristic	IM	Other HQIM	Effect size difference
Math achievement (z-score)	0.41	0.41	0.00
Students from households with low income	32%	31%	0.03
Students with disabilities	19%	19%	0.11
Sample size	169	169	-

Note. Effect size is expressed in standardized units using Cohen's *d*.

For the matched non-HQIM group, however, there were large differences in prior achievement trends even after matching. Accordingly, we could not be confident that post-IM differences were due to IM or due to these preexisting differences. Accordingly, in this main report, we highlight only the results using the other HQIM grade–school observations as the comparison group.

The findings first reflect simple aggregated mean differences in math achievement in the first 3 years following IM implementation between IM grade-school observations and their matched HQIM counterparts, accompanied with statistical significance (via *t*-tests) and standardized effect sizes. Since the difference in students with disabilities was greater than 0.05 standard deviations, we controlled for this characteristic, along with prior math achievement and students from households with low income, in the regression model used to estimate IM impact. The resulting adjusted IM impact estimate from this model is provided, by year and cohort, in Table A2 below.

Table A2. Regression-Adjusted Estimates of IM Impact Versus Matched HQIM Observations

	Year 1	Year 2	Year 3
All observations	0.03	0.05	0.19***
Pre-COVID-19 cohort	0.18***	-	0.39***
Post-COVID-19 cohort	-0.04	-0.11	-0.06

Note. Cell values indicate adjusted difference between IM and matched HQIM observations in standard deviation units. Year 2 estimates were only available for the post-COVID cohort because spring 2020 corresponded to Year 2 for the pre-COVID cohort, when assessments were not administered. Asterisks indicate statistical significance: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

As described in the report, the adjusted impact of IM was driven entirely by the pre-COVID-19 cohort. The adjusted impact was 0.18 standard deviations in Year 1 and 0.39 standard deviations in Year 3.