ORIGO
STEPPING STONES
CORE MATHEMATICS PROGRAM

The Effect of Stepping Stones Math on Student Achievement:

An Independent Evaluation of the *Stepping Stones* Math Curriculum by American Institutes for Research® (AIR)

Executive Summary

The American Institutes for Research (AIR) was contracted by ORIGO Education to conduct an external independent evaluation of the *Stepping Stones* Math curriculum. The study was designed to estimate the effect of *Stepping Stones* on student mathematics achievement, as measured by the percentage of students who scored proficient or better on state end-of-year assessments.

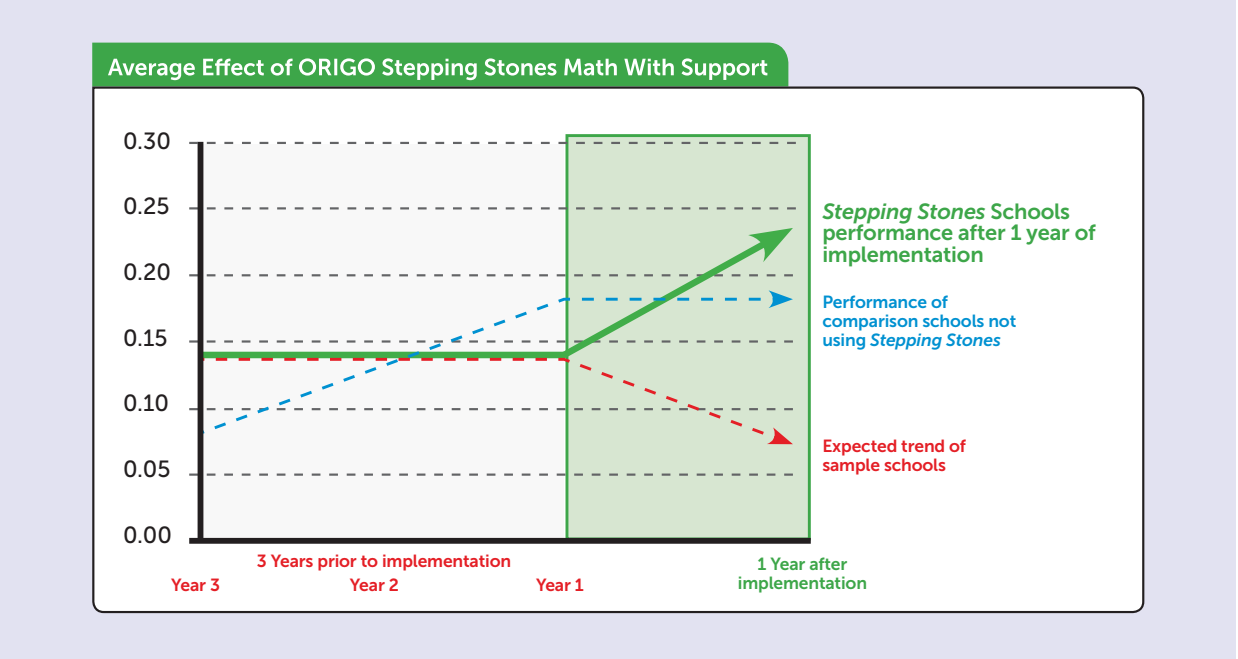
The two key research questions were:

1. To what extent does access to *Stepping Stones* affect students' mathematics achievement in the first year?
2. To what extent does access to *Stepping Stones* affect students' mathematics achievement in the first year in districts that purchased professional learning for *Stepping Stones*?

The study sampled schools had access to *Stepping Stones* beginning in the 2014-15, 2015-16, 2016-17, or 2017-18 school year for at least one grade between Grades 3 to 5. This totaled 559 grade-level classes across 202 schools and 39 districts in six states: California, Illinois, Kentucky, Ohio, Washington, and Wisconsin. A grade-level class includes all the students in a given grade in one building, regardless of how many classroom teachers serve that grade in that building.

In districts that invested in a meaningful amount of professional learning (at least 10 days for the districts in this study) to support high-quality implementation of *Stepping Stones*, **the estimated change in state mathematics performance was about 0.14 standard deviations better than one would expect—a statistically significant change.** An effect of 0.14 standard deviations suggests that **an average class in the state would be ranked at approximately the 56th percentile instead of the 50th percentile if it had access to *Stepping Stones* – a gain of 6 percentile points.**

The analysis was designed to meet the What Works Clearinghouse group design standards with reservations (Tier 2), which is the highest possible rating for a quasi-experimental design.



The study found that pairing *Stepping Stones* Math with sufficient implementation support has a significant positive effect on mathematics proficiency. The study design measured the percent of students scoring at least proficient on state assessments, a less sensitive measure than average scale score. The study design also assumes that each teacher implements the program with equal effectiveness. A future study could show the impact of *Stepping Stones* to be greater if:

- the study used student scale scores as the measure of achievement because those are more sensitive to gains in individual learning; and
- the study was able to control for the effectiveness of *Stepping Stones* implementation because we would expect a greater impact in classrooms where the program is implemented more effectively.

A more precise read on the effectiveness of implementation can be obtained with high quality tools. **During the time of this study, ORIGO Education developed tools for supporting full implementation of *Stepping Stones*. The result is the *Stepping Stones* Implementation Tool (SSIT), designed to provide District and/or School Implementation Teams with an efficient measure of the extent to which school personnel are applying the core elements of the *Stepping Stones* program in classrooms.** This tool is intended to be used over time to guide implementation planning of the *Stepping Stones* program, not to evaluate teacher performance. **The *Stepping Stones* Implementation Tool supports schools with reaching the desired level of implementation and resulting impact on student achievement described by this report.**

ORIGO

STEPPING STONES 2.0

CORE MATHEMATICS PROGRAM

A core mathematics program designed for conceptual understanding.

ORIGO Stepping Stones 2.0 guides Pre-K through Grade 6 students to proficiency with college and career readiness standards.

Developed by a team of dedicated experts working with educators like you, *Stepping Stones* is designed to close the achievement gap, support at-risk learners, and inspire all students toward future success.

Stepping Stones integrates a unique set of digital tools and print supplements, as well as embedded professional learning to give your students and teachers a flexible, balanced, and well supported experience.

We shift learning from rote memorization to deep conceptual understanding.

Teach with confidence using visual models and language, engaging applications, and rich assessment options — all from the start.

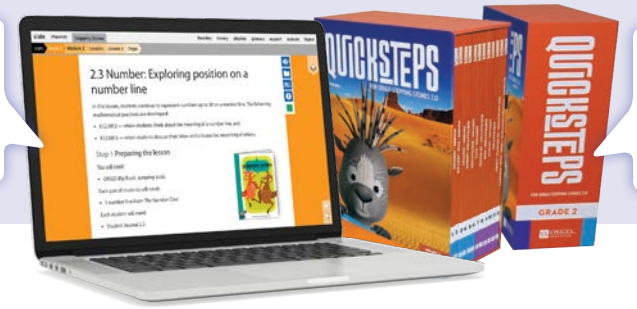
ORIGO Stepping Stones helps close the achievement gap for English language learners, special education, and economically disadvantaged students.

Grades K–6

Stepping Stones Program Components

Teacher Edition

Digital Teacher Edition



Printed Teacher Edition

Student Journal

Digital Student Journals

Printed Student Journals

Available in English and Spanish

Online Assessment

Assessment

Reporting

Resources

ORIGO Big Books

12 books per grade, available in English and Spanish

The Number Case

Available in English and Spanish

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

The Effect of Stepping Stones Math on Student Achievement

FEBRUARY 2020

Jordan Rickles | So Jung Park

MAKING RESEARCH **RELEVANT**

The Effect of Stepping Stones Math on Student Achievement

FEBRUARY 2020

Jordan Rickles | So Jung Park

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Contents

Exhibits

Introduction

The American Institutes for Research (AIR) was contracted by ORIGO Education to conduct an external independent evaluation of the Stepping Stones Math curriculum. Stepping Stones is a comprehensive mathematics curriculum for Grades K–5 that is aligned to the Common Core State Standards. Based online, Stepping Stones provides access to a set of digital tools and print supplements. In addition, districts can purchase teacher professional learning sessions from ORIGO Education to support implementation of Stepping Stones. This study evaluates the effect of Stepping Stones on mathematics achievement by comparing schools with access to Stepping Stones to schools without Stepping Stones. Data for this study come from six states: California, Illinois, Kentucky, Ohio, Washington, and Wisconsin.

This report presents the study findings for the following research questions:

1. To what extent does access to Stepping Stones affect students' mathematics achievement in the first year of access?
2. In schools with access to Stepping Stones in multiple years, to what extent does access to Stepping Stones affect students' mathematics achievement over 3 years?
3. To what extent does access to Stepping Stones affect student mathematics achievement in the first year in different school settings?
    a. What is the average effect of Stepping Stones in its first year of implementation in schools in which most students come from low income backgrounds?
    b. What is the average effect of Stepping Stones in its first year of implementation in schools in which many students are an underrepresented minority?
4. To what extent does access to Stepping Stones affect student mathematics achievement in the first year in districts that purchased professional learning for Stepping Stones?
    a. What is the average effect of Stepping Stones in its first year of implementation in districts in which at least 1 day of professional learning was purchased?
    b. What is the average effect of Stepping Stones in its first year of implementation in schools in which at least 10 days of professional learning was purchased?[1]

The results suggest that, on average, Stepping Stones did not have a significantly positive effect on student mathematics achievement among all schools that had access to the program. But there is a significant positive effect among districts that purchased Stepping Stones and at least 10 days of professional learning (see Key Findings box).

[1] Based on the distribution of professional learning days purchased among the districts in the study, 10 days of professional learning represented a natural proxy for districts that made a substantial investment in supporting implementation of Stepping Stones.

Key Findings

- In classes that had access to Stepping Stones, the estimated change in state mathematics performance in the first year of access was about 0.07 standard deviations better than one would expect from similar schools without Stepping Stones, a change that was not statistically significant.
- In classes that had access to Stepping Stones for 3 years, the estimated effect was similar in each of the 3 years and not statistically different from zero (0.07 standard deviations in Year 1, 0.06 standard deviations in Year 2, and 0.05 standard deviations in Year 3), implying that just having access to Stepping Stones does not significantly improve mathematics performance over time.
- The estimated effect of having access to Stepping Stones during the first year was relatively similar across school settings, specifically in lower income schools (0.06 standard deviations) and higher minority schools (0.09 standard deviations). These effect estimates were not statistically significant.
- In districts that invested in a meaningful amount of professional learning (at least 10 days for the districts in this study) to support high-quality implementation of Stepping Stones, the change in relative state mathematics performance in the first year suggests that Stepping Stones significantly improved average mathematics performance by 0.14 standard deviations. This change implies that a class at the 50th percentile in the state would be at approximately the 56th percentile if it had access to Stepping Stones and professional learning support.

About the Analysis in This Report

This report focuses on the effectiveness of Stepping Stones for schools that had access to Stepping Stones in the 2014–15, 2015–16, 2016–17, or 2017–18 school year for at least one grade between Grades 3 and 5.

Data Sources

AIR used three data sources for the study: school enrollment and characteristics from the Common Core of Data (CCD),[2] aggregate public-use data from end-of-year mathematics assessments in each state, and Stepping Stones Math curriculum purchasing data from ORIGO Education.

Study Sample

The analysis is based on the grade-level classes (defined as grades within schools) that had access to Stepping Stones in six states: California, Illinois, Kentucky, Ohio, Washington, and Wisconsin. We do not have assessment data on individual students or classrooms within a school; therefore, our unit of analysis is the grade-level class, in aggregate, within schools for a particular year—for example, Grade 4 in Springfield Elementary School for the 2015–16 school year. For simplicity, we refer to this school-by-grade-level unit of analysis as *classes*.

[2] We used school enrollment and characteristics from the 2014–15 CCD. A small subset of schools ($N = 4$) in our analytic sample was not in the 2014–15 CCD. For those schools, we used the 2015–16 or 2016–17 CCD.

Classes were defined as being in one of four treatment cohorts based on the year in which they first had access to Stepping Stones: Cohort 1 (2014–15), Cohort 2 (2015–16), Cohort 3 (2016–17), and Cohort 4 (2017–18). To be included in the analysis, classes had to have state mathematics assessment results in the state's publicly available reporting system for 3 years before the start of Stepping Stones (the cohort start year) and up to 3 years of data after the cohort start year through the 2017–18 school year.[3] We required assessment data over this period to produce a reliable estimate of achievement before the cohort start year (baseline period), and to estimate the effect of Stepping Stones over multiple years. The years of required data for each cohort are depicted in Exhibit 1.

Exhibit 1. Years of Required Data by Cohort

	School Year						
Cohort	**2011–12**	**2012–13**	**2013–14**	**2014–15**	**2015–16**	**2016–17**	**2017–18**
Cohort 1 (2014–15)	Baseline Year 3	Baseline Year 2	Baseline Year 1	**Cohort Start Year**	Follow-Up Year 2	Follow-Up Year 3	
Cohort 2 (2015–16)		Baseline Year 3	Baseline Year 2	Baseline Year 1	**Cohort Start Year**	Follow-Up Year 2	Follow-Up Year 3
Cohort 3 (2016–17)			Baseline Year 3	Baseline Year 2	Baseline Year 1	**Cohort Start Year**	Follow-Up Year 2
Cohort 4 (2017–18)				Baseline Year 3	Baseline Year 2	Baseline Year 1	**Cohort Start Year**

Overall, of the 631 classes that had access to Stepping Stones, 559 classes across 202 schools and 39 districts met the analysis criteria for inclusion (See Exhibit 2). Of the classes that met the inclusion criteria, 34% were Grade 3, 35% were Grade 4, and 31% were Grade 5. The number of Stepping Stones classes by state and cohort is provided in Appendix A (Exhibit A.1).

[3] An exception to this inclusion rule was made for Wisconsin because state assessments were not officially administered in 2014–15, and no test results are publicly available. For Cohorts 2 through 4 in Wisconsin, we did not require data from the 2014–15 school year, but we did require data from one additional year before the cohort start year (2015–16 for Cohort 2, 2016–17 for Cohort 3, and 2017–18 for Cohort 4). This requirement ensures that analyses for all schools are based on at least 3 years of prior test data. We did not include Wisconsin's Cohort 1 (90 classes) in the analysis of the first research question because there are no test results for the cohort start year (2014–15).

Exhibit 2. Number of Districts, Schools, and Grade-Level Classes That Met the Analysis Inclusion Criteria

State	Districts	Schools	Grade-Level Classes
CA	3	44	132
IL	9	28	76
KY	6	18	53
OH	4	18	40
WA	6	29	77
WI	11	65	181
Total	**39**	**202**	**559**

Comparison classes were selected through a combination of exact matching and propensity score matching.[4] Exact matching, in which comparison classes are selected only if they have the exact same school characteristics as their paired Stepping Stones class, was defined by the following characteristics: state, grade level, locale (city, suburban, town/rural), magnet school, and charter school. Within each exact match category, a single comparison class was matched to each Stepping Stones class using a propensity score estimated with the following characteristics: Title I status, total enrollment, percent free or reduced-price lunch, race/ethnicity composition, and mathematics achievement during the 3 years before using Stepping Stones. Overall, we were able to find matches for 554 Stepping Stones classes (97% match rate). For these 554 classes, the matching process resulted in one comparison class for every Stepping Stones class included in an analysis (i.e., a one-to-one match without replacement), so that the total sample size for a specific analysis is twice the number of Stepping Stones classes included in the analysis.

Using the matched sample, we created analytic samples for each research question. Exhibit 3 shows the number of Stepping Stones classes for each analysis.

- For research question (RQ) 1, which focuses on the effect of Stepping Stones in the first year of access, we included all schools in the sample that started using Stepping Stones between 2014–15 and 2017–18 and have 3 years of prior testing data. For the RQ 1 analysis, we could not include 90 Cohort 1 classes in Wisconsin (and their matched comparison classes) because Wisconsin does not have 2014–15 achievement data, which is the first-year outcome for Cohort 1.

[4] A propensity score is the predicted probability that a class had access to Stepping Stones based on observed characteristics of the class. Classes with similar propensity scores have similar characteristics, and matching classes based on their estimated propensity score is an efficient way to match classes across multiple characteristics.

- For RQ 2, which focuses on the effect of Stepping Stones over 3 years, we looked only at schools that started using Stepping Stones in 2014–15 (Cohort 1) or 2015–16 (Cohort 2) because we have at least 3 years of outcome data for these schools. This analysis lets us examine whether the effect of Stepping Stones increased or decreased after the first year for the classes that had access over 3 years.
- The analysis for RQ 3 and RQ 4 mirrors the RQ 1 analysis but focuses on specific types of schools and districts to examine whether the effect of the Stepping Stones curriculum differs by school and district characteristics. For RQ 3a, we restricted the analysis to "lower income" schools, defined as schools with at least half of the students eligible for free or reduced-price lunch. For RQ 3b, we restricted the analysis to "higher minority" schools, defined as schools with at least a third of the students from a non-White racial/ethnic group. Schools were included in the RQ 3 analysis only if the Stepping Stones school and its matched comparison school had the defining characteristic.[5]
- For RQ 4a, we restricted the analysis to Stepping Stones schools in districts that purchased at least 1 day of professional learning and their matched comparison schools. For RQ 4b, we restricted the analysis to Stepping Stones schools in districts that purchased at least 10 days of professional learning and their matched comparison schools.

Appendix A (Exhibit A.3) provides information on how similar, on average, the Stepping Stones classes are to the matched comparison group. A general rule of thumb for determining whether the two groups are similar is if the standardized mean difference on measurable characteristics is less than 0.25 standard deviations. This is the case for all the characteristics reported in Exhibit A.3. The Technical Supplement provides parallel information about how similar the Stepping Stones classes and matched comparison classes are for each research question. For all research questions, the two groups have similar mathematics performance before the introduction of Stepping Stones, which is an important condition for making valid comparisons between the two groups.

[5] We only included states in the RQ 3a and RQ 3b analysis if they had at least 5 Stepping Stones grade-level classes in the school type of interest. Ohio had only 3 classes in lower income schools and Kentucky had only 3 classes in higher minority schools.

Exhibit 3. Number of Stepping Stones Grade-Level Classes Included in Each Analysis, by State

State	RQ 1: All Grade-Level Classes	RQ 2: Grade-Level Classes With 3 Years of Access	RQ 3a: Grade-Level Classes in Lower Income Schools	RQ 3b: Grade-Level Classes in Higher Minority Schools	RQ 4a: Grade-Level Classes in Districts With Any PL	RQ 4b: Grade-Level Classes in Districts With at Least 10 Days of PL
CA	132	129	30	123	123	123
IL	76	65	22	40	65	15
KY	53	53	36	0	18	0
OH	40	7	0	8	31	31
WA	77	42	34	43	70	38
WI	76	151	34	41	49	0
Total	**454**	**447**	**156**	**255**	**356**	**207**

Note. PL = professional learning.

Measure of Mathematics Achievement

We used the mathematics percent proficient on the state end-of-year assessment for each class as our measure of mathematics achievement. This is because only two of the six states had public-use assessment data that included average mathematics scale scores for the study time period, while the public-use assessment data for all six states included the percentage of students who scored at or above the state's proficiency benchmark. To properly examine differences in percent proficient across classes and time, we converted each percent proficient score to a log-odds score and then standardized the log-odds score (see the Technical Supplement for more details). This standardization puts all the percent proficient scores on a common scale (*z*-scores), making it possible to draw valid comparisons when tests (or proficiency standards) change over time or differ across states.[6] The assessment systems used by each state during the study period are reported in Appendix A (Exhibit A.2).

Analytic Approach

We used a comparative interrupted time series (CITS) design to estimate the effect of Stepping Stones on student mathematics achievement. The CITS design compares baseline period to

[6] The *z*-score conversion was based on the log-odds mean and standard deviation for a particular grade, state, and year. Means and standard deviations were calculated with the grade-level, public-use assessment data; thus, a *z*-score of zero represents the score in a particular grade and year for the average class, and a *z*-score of one represents the score in a particular grade and year that is one standard deviation higher than the average class.

follow-up period changes in average mathematics achievement in Stepping Stones classes to baseline period to follow-up period changes in similar classes that did not have access to the Stepping Stones Math curriculum. The analysis was designed to meet the What Works Clearinghouse group design standards with reservations, which is the highest possible rating for a quasi-experimental design. CITS designs are considered one of the most rigorous quasi-experimental designs because they reduce the potential for internal validity threats (e.g., history and selection bias) by using historical information about outcome trajectories among the treatment units as well as information about outcome trajectories among comparison units.

For the analysis of Stepping Stones effectiveness, we used a baseline trend model. Using the baseline trend model, program impacts are evaluated by looking at whether the treatment group deviates from its baseline trend by a greater amount than the comparison group's deviation from its trend. The matching process identified comparison classes that are similar to the Stepping Stones classes in terms of prior assessment scores and school characteristics. To further account for some group differences that exist after matching, we controlled for each class's grade, cohort, and school characteristics in the statistical model. This approach accounts for differences in the achievement trajectories between the Stepping Stones and comparison classes and can provide valid estimates of the Stepping Stones effect on student achievement as long as any factors associated with changes in student achievement, such as changes in state testing, affect the Stepping Stones and comparison schools equally.

We estimated the effect of Stepping Stones separately for each state. To summarize results across states, we report weighted averages of the state-specific effect estimates.

More details about the modeling approach are provided in the Technical Supplement.

Findings

In this section, we present findings for each research question based on the six-state weighted average effect estimates. State-specific results are presented in Appendix B.

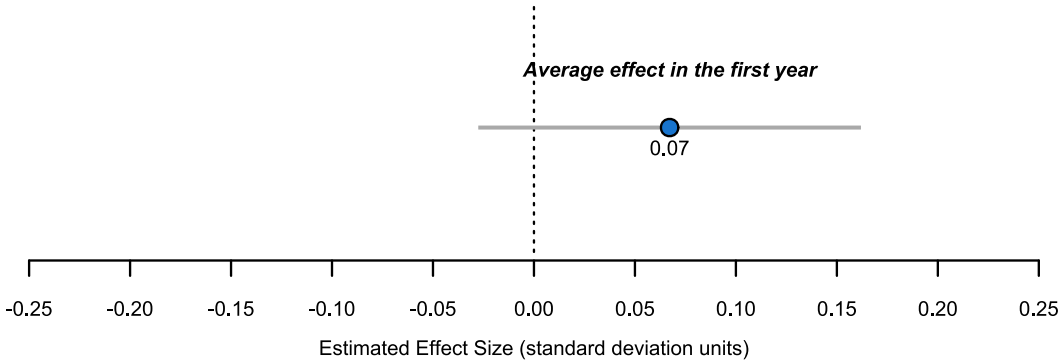
RQ 1: The Effect of Stepping Stones Access in the First Year

The analysis of classes in the six states suggests that access to Stepping Stones did not have a statistically significant effect on mathematics achievement. On average, classes with access to Stepping Stones performed about 0.07 standard deviations better relative to the comparison classes. This effect size is depicted in Exhibit 4. The exhibit shows the estimated average effect (circle) and the 95% confidence interval for the effect (horizontal gray line). Effect estimates with a confidence interval that does not intersect with zero are statistically significant at the conventional 0.05 alpha level. Because the confidence interval crosses zero, we conclude that

there is too much uncertainty in the effect estimate to rule out the possibility that Stepping Stones had no effect on mathematics achievement.

In addition, the estimated average effect, while positive, is small in magnitude.[7] An effect of 0.07 standard deviations suggests that an average class in the state would be ranked at approximately the 53rd percentile instead of the 50th percentile if it had access to Stepping Stones.

Exhibit 4. Estimated Effect of Access to Stepping Stones on Mathematics Achievement in the First Year of Access



Notes. The exhibit presents the estimated effect (circle) of having access to Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray line represents the 95% confidence interval for the estimated effect. *N* = 908 grade-level classes (Stepping Stones *N* = 454; Comparison *N* = 454).

RQ 2: The Effect of Stepping Stones Access Across 3 Years

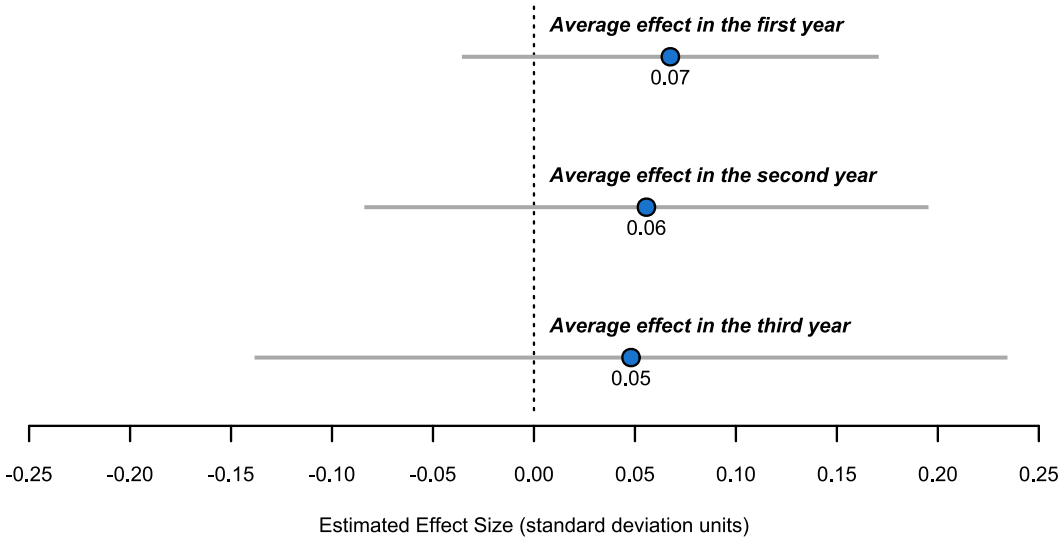
Schools may get better at implementing Stepping Stones over time. If so, looking at the effect of Stepping Stones over multiple years may provide a more general indication of the program's effectiveness than looking at just the first-year results. For this analysis, we focused on classes with at least 3 years of access to Stepping Stones.[8]

Results based on 3 years of mathematics achievement after classes had access to Stepping Stones indicate that the average effect of Stepping Stones was relatively stable across the 3 years (see Exhibit 5). As with our primary analysis of the effect in the first year, the effect in each of the 3 years with access are small in size and not statistically significant.

[7] Because the effect sizes reported for this study are based on the standard deviation among class means, rather than student-level scores, they are likely larger than effect sizes that one would get with student-level scores. One should not directly compare the magnitude of the effects in this report to the magnitude of effects from studies of student-level data. For example, a class-level effect size of 0.07 would be an effect size of about 0.02 or 0.03 depending on how much student achievement varies within classes.

[8] Of the classes in the RQ 2 analysis, 80% were in the RQ 1 analysis. Characteristics of the RQ 1 and RQ 2 samples are provided in the Technical Supplement.

Exhibit 5. Estimated Effect of Access to Stepping Stones on Mathematics Achievement Over 3 Years of Access



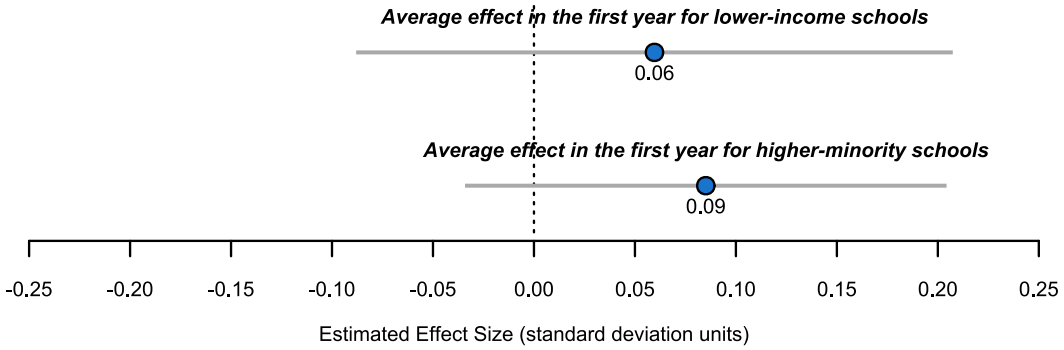
Notes. The exhibit presents the estimated effect (circles) of having access to Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect.
N = 894 grade-level classes (Stepping Stones *N* = 447; Comparison *N* = 447).

RQ 3: The Effect of Stepping Stones Access in Different School Settings

The results for RQ 1 and RQ 2 focus on the average effect of Stepping Stones across schools in the six states, but the effect may differ across school settings. To explore this possibility and gain a better understanding of whether the overall average effects generalize to different school settings, we looked at the effect of Stepping Stones access in lower income schools (defined as schools with at least half of their students eligible for free or reduced-price lunch) and higher minority schools (defined as schools with at least a third of their students classified as a non-White racial/ethnic group). We focus on the first-year effect for this analysis.

The estimated effect size for each school type is reported in Exhibit 6. The estimated average effect for lower income schools is, like the overall effect presented in Exhibit 4, small in magnitude and not statistically significant. For schools serving a higher percentage of minority students, the estimated effect is slightly higher (0.09 standard deviations) but not statistically different from zero.

Exhibit 6. Estimated Effect of Access to Stepping Stones on Mathematics Achievement in the First Year of Access, by School Setting



Notes. The exhibit presents the estimated effect (circles) of having access to Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect.
N = 312 lower- income grade-level classes (Stepping Stones *N* = 156; Comparison *N* = 156); 510 higher- minority grade-level classes (Stepping Stones *N* = 255; Comparison *N* = 255).

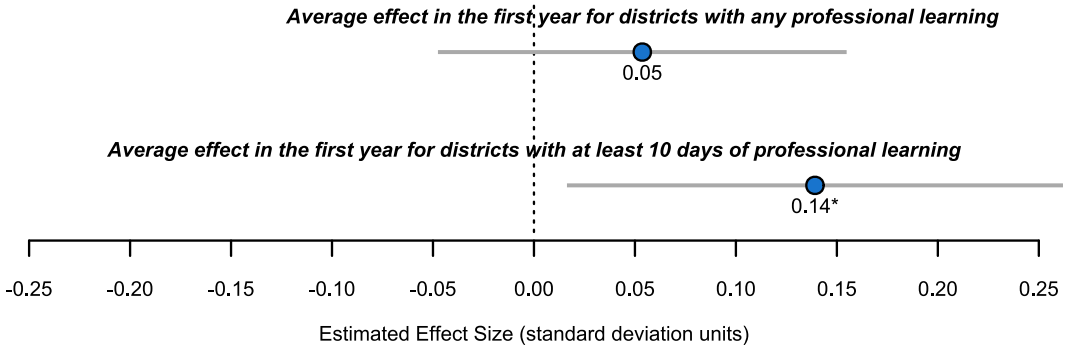
RQ 4: The Effect of Stepping Stones Access in Districts That Provided Professional Learning Opportunities

The results for the first three research questions are based on all schools with access to Stepping Stones, but not all schools received the same level of training on how to incorporate Stepping Stones into their mathematics instruction. While ORIGO Education does not have detailed data on which schools received training, it does have information on which districts purchased PL for their schools. To examine the effect of Stepping Stones when schools have access to training, we estimated the effect using two nested samples. The first sample includes districts that purchased at least 1 day of PL, representing a liberal definition of access to training. The second sample is restricted to districts that purchased at least 10 days of professional learning to focus on classes that had a better opportunity to receive enough training to sufficiently support implementation with fidelity. The threshold of 10 PL days represents a useful proxy for districts that made a substantial investment in supporting implementation of Stepping Stones, but the actual number of days required to sufficiently support implementation could differ based on district size and school capacity. Of classes with access to any professional learning, 58% were in districts that purchased at least 10 days of PL. We focus on the first-year effect for this analysis.

The estimated effect size for each sample is reported in Exhibit 7. The estimated average effect in districts with any professional learning purchased is, like the overall effect presented in Exhibit 4, modest in magnitude and not statistically significant. For classes in districts that provided at least

10 days of PL, however, the estimated effect size is about twice the size of the effect for the total sample (0.14 standard deviations) and is statistically significant. An effect of 0.14 standard deviations suggests that an average class in the state would be ranked at approximately the 56th percentile instead of the 50th percentile if it had access to Stepping Stones.

Exhibit 7. Estimated Effect of Access to Stepping Stones on Mathematics Achievement in the First Year of Access, by Number of Professional Learning Days Purchased by the District



Notes. The exhibit presents the estimated effect (circles) of having access to Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect.
N = 712 grade-level classes in the any professional learning sample (Stepping Stones *N* = 356; Comparison *N* = 356); 414 grade-level classes in at least 10 days of professional learning sample (Stepping Stones *N* = 207; Comparison *N* = 207).
* The estimated effect is statistically significant at the 0.05 alpha level (two-tailed test).

Study Limitations

Two limitations of the available data may hinder our ability to detect a meaningful effect of Stepping Stones. First, the available data from ORIGO Education identified which schools had a license to use Stepping Stones but did not include more detailed information that would enable us to identify how many students actually used Stepping Stones or how well it was implemented in specific classrooms. As a result, our sample of Stepping Stones classes may include classes with little or no exposure to the Stepping Stones program. Including non-users in the sample of Stepping Stones classes likely waters down the estimated effect of the Stepping Stones curriculum.

Second, our analysis of mathematics achievement was limited to grade-level data on the percentage of students who scored proficient or better on the state test. These data are likely

less sensitive to changes in achievement than average scale scores.[9] As a result, the estimated effect of Stepping Stones may not reflect changes concentrated among the lowest performing students (who may improve but not achieve proficiency) or the highest performing students (who may improve but remain proficient regardless).

In addition, the quasi-experimental nature of this study means we cannot rule out the possibility that mathematics achievement differences between Stepping Stones schools and comparison schools were due to other factors besides Stepping Stones. While we used a rigorous quasi-experimental design to guard against common biases, the results could be sensitive to misspecifications of the analytic model or to unobserved factors that coincided with the adoption of Stepping Stones and affected Stepping Stones schools differently than the comparison schools. As such, additional research of Stepping Stones is required to make more definitive statements about its effectiveness.

Conclusion

Overall, the results suggest that having access to Stepping Stones does not have a significant effect on mathematics proficiency. However, given the positive direction of the estimated effects and the study limitations discussed above, one should not rule out the possibility that Stepping Stones has a positive effect on student learning when well implemented. This point is highlighted by the finding that Stepping Stones had a small statistically significant positive effect in districts that purchased Stepping Stones and at least 10 days of professional learning. While 10 professional learning days served as a useful marker for sufficient implementation support in our analysis, the actual number of professional learning days for effective implementation of Stepping Stones could differ based on district size and school capacity.

[9] We had to use the "percent proficient" measure because average scale scores were not consistently reported in the state public-use data files.

Appendix A. Supporting Exhibits

Exhibit A.1. Number of Grade-Level Classes With Access to Stepping Stones, by State and Cohort

	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Total	
State	Total	Eligible	Total	Eligible	Total	Eligible	Total	Eligible	Total	Eligible
CA	51	51	78	78	3	3	0	0	132	132
IL	24	23	42	42	11	11	0	0	77	76
KY	42	38	15	15	0	0	0	0	57	53
OH	15	7	0	0	33	33	0	0	48	40
WA	42	41	1	1	20	20	15	15	78	77
WI	96	96	70	70	15	15	0	0	181	181
Total	**270**	**256**	**206**	**206**	**82**	**82**	**15**	**15**	**573**	**559**

Exhibit A.2. Assessment Systems, by State and Year

	School Year						
State	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17	2017–18
CA	CST	CST	–	SB	SB	SB	SB
IL	ISAT	ISAT	ISAT	PARCC	PARCC	PARCC	PARCC
KY	KPREP	KPREP	KPREP	KPREP	KPREP	KPREP	KPREP
OH	OAA	OAA	OAA	PARCC	PARCC	PARCC	PARCC
WA	MSP	MSP	MSP	SB	SB	SB	SB
WI	WSAS	WSAS	WSAS	–	FORWARD	FORWARD	FORWARD

Notes.
CST = California Standards Tests
ISAT = Illinois Standards Achievement Test
KPREP = Kentucky Performance Rating for Educational Progress
MSP = Measurements of Student Progress (WA)
OAA = Ohio Achievement Assessments
PARCC = Partnership for Assessment of Readiness for College and Careers
SB = Smarter Balanced
WSAS = Wisconsin Student Assessment System

Exhibit A.3. Characteristics of Matched Stepping Stones and Comparison Grade-Level Classes Pooled Across Cohorts and States

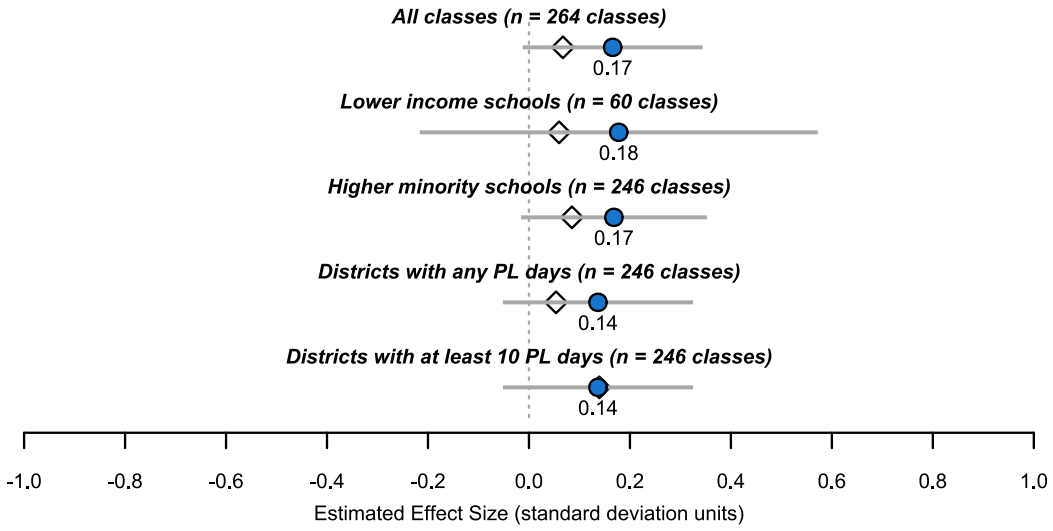
Characteristic	Matched Grade-Level Classes			All Eligible Comparison Grade-Level Classes	
	SS Mean (*N* = 544)	C Mean (*N* = 544)	SMD	C Mean (*N* = 117,910)	SMD
Location: City	0.43	0.43	0.00	0.36	0.17
Location: Suburb	0.39	0.39	0.00	0.39	0.01
Location: Town/Rural	0.18	0.18	0.00	0.25	-0.26
Magnet School	0.02	0.02	0.00	0.03	-0.37
Charter School	0.02	0.02	0.00	0.04	-0.60
Title I: Targeted	0.19	0.18	0.01	0.16	0.10
Title I: Schoolwide	0.58	0.56	0.05	0.64	-0.14
Total Enrollment (Grades 3–8)	483.28	514.94	-0.18	514.57	-0.13
Percent NSLP	0.47	0.46	0.06	0.59	-0.39
Percent Black	0.06	0.05	0.10	0.10	-0.21
Percent Asian/Pacific Islander	0.07	0.06	0.09	0.07	0.06
Percent Hispanic	0.27	0.28	-0.02	0.35	-0.24
Percent White	0.54	0.55	-0.06	0.43	0.30
Percent Proficient: Prior Year 1 (*z*-score)	0.10	0.17	-0.07	0.00	0.10
Percent Proficient: Prior Year 2 (*z*-score)	0.07	0.16	-0.08	0.00	0.07
Percent Proficient: Prior Year 3 (*z*-score)	0.03	0.08	-0.06	0.00	0.03

Notes. The statistics for all eligible comparison grade-level classes provide a general reference for how the Stepping Stones grade-level classes compare to the broader population of eligible grade-level classes and how matching improved comparability. The means for the four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table. The Technical Supplement includes exhibits that report the characteristics of the matched Stepping Stones and comparison grade-level classes for each state. A general rule of thumb for determining whether the two groups are similar is if the standardized mean difference on measurable characteristics is less than 0.25 standard deviations.
SS = Stepping Stones class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

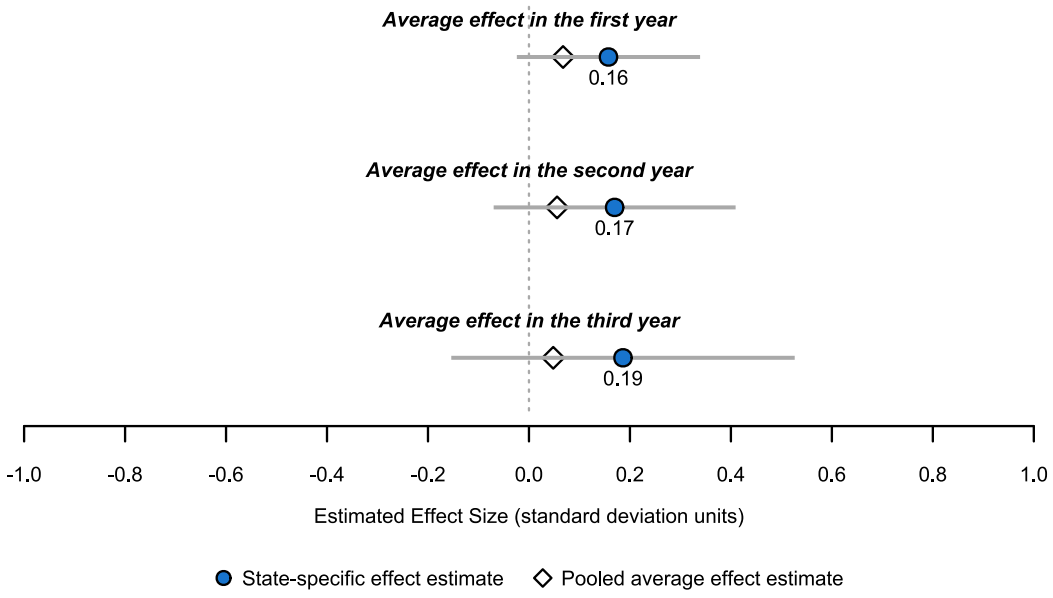
Appendix B. State-Specific Results

Exhibit B.1. Estimated Effects in California

A. Effects in the first year of access



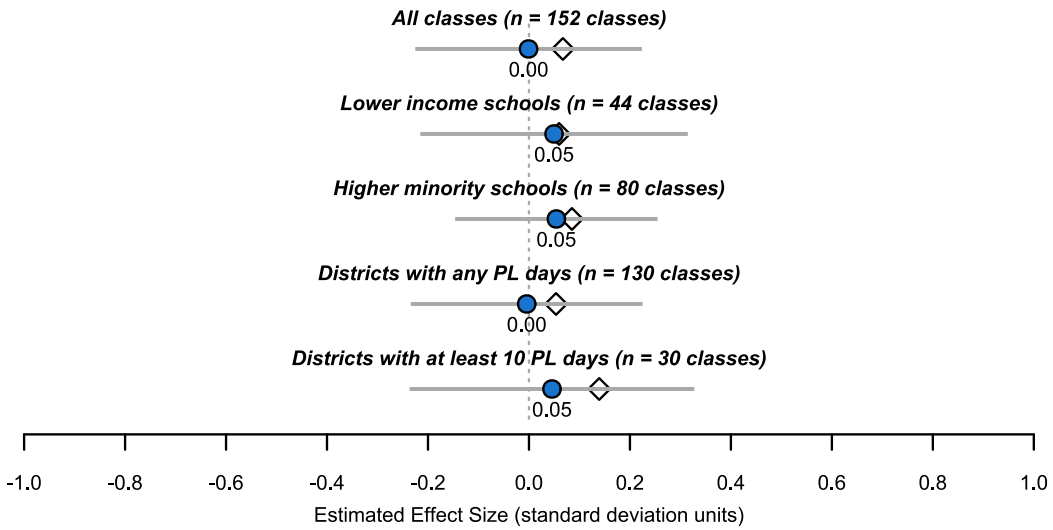
B. Effects over 3 years of access in grade-level classes with 3 years of access (*n* = 258 grade-level classes)



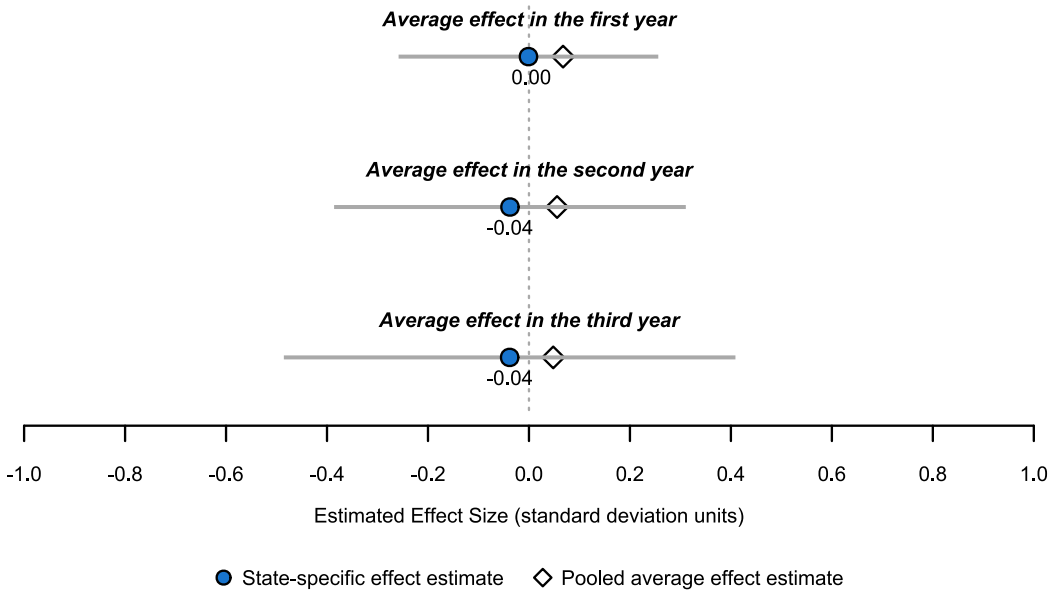
Notes. The exhibit presents the estimated effect (circles) of Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect. The diamonds depict the average effect across all states in the study.

Exhibit B.2. Estimated Effects in Illinois

A. Effects in the first year of access



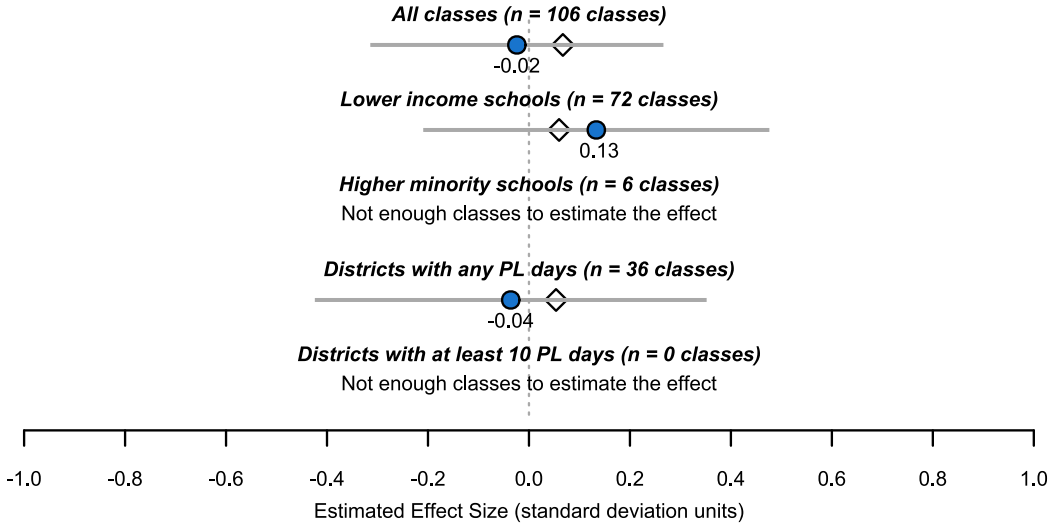
B. Effects over 3 years of access in grade-level classes with 3 years of access (*n* = 130 grade-level classes)



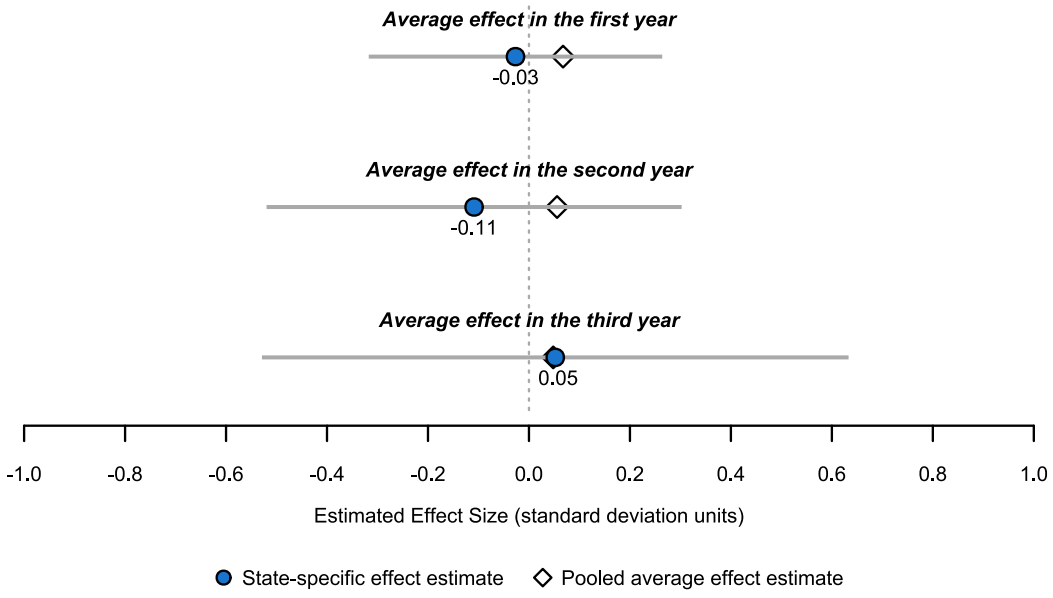
Notes. The exhibit presents the estimated effect (circles) of Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect. The diamonds depict the average effect across all states in the study.

Exhibit B.3. Estimated Effects in Kentucky

A. Effects in the first year of access



B. Effects over 3 years of access in grade-level classes with 3 years of access (*n* = 106 grade-level classes)

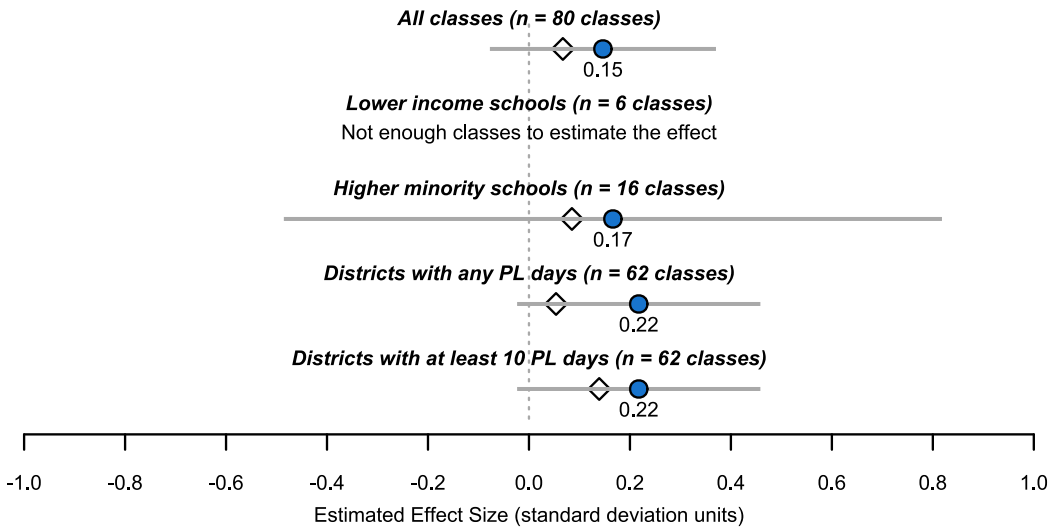


Notes. The exhibit presents the estimated effect (circles) of Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect. The diamonds depict the average effect across all states in the study.

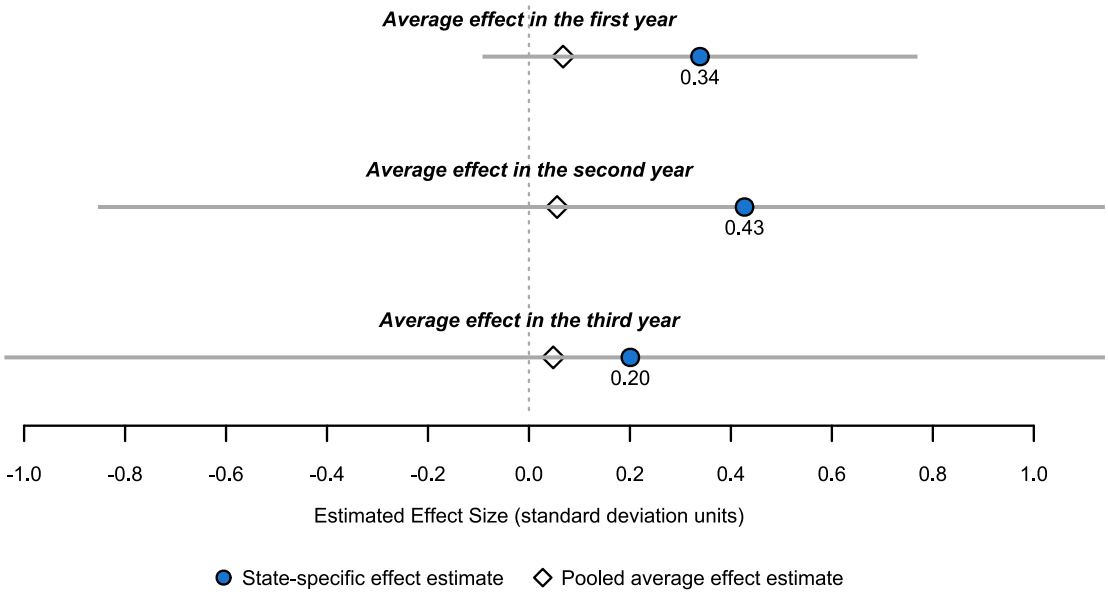
* The estimated effect is statistically significant at the 0.05 alpha level (two-tailed test).

Exhibit B.4. Estimated Effects in Ohio

A. Effects in the first year of access



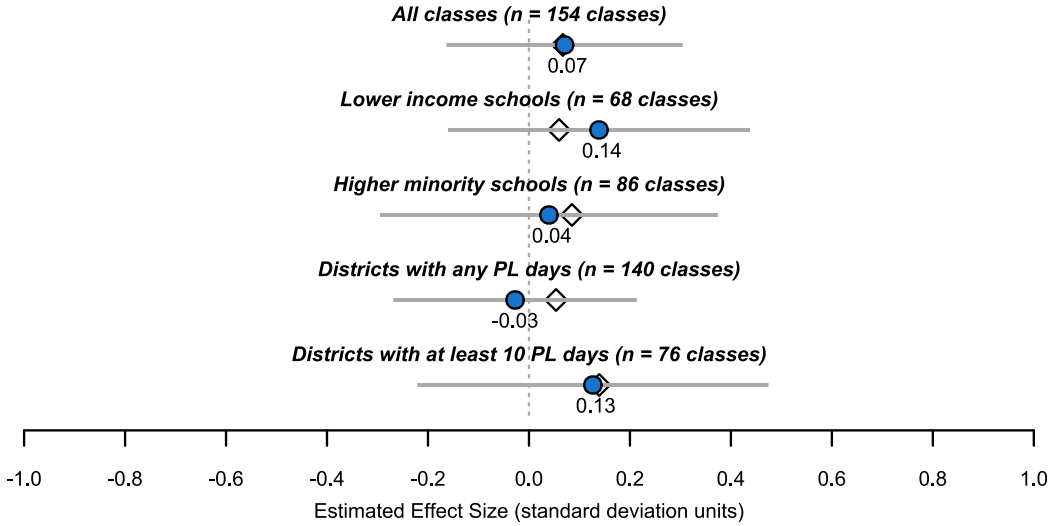
B. Effects over 3 years of access in grade-level classes with 3 years of access (*n* = 14 grade-level classes)



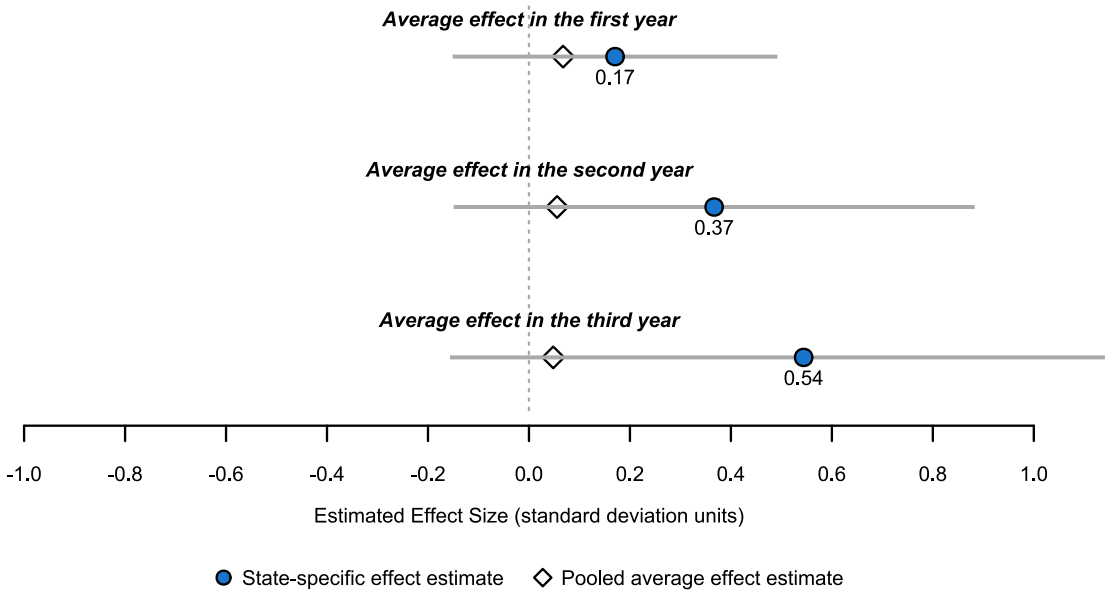
Notes. The exhibit presents the estimated effect (circles) of Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect. The diamonds depict the average effect across all states in the study.

Exhibit B.5. Estimated Effects in Washington

A. Effects in the first year of access



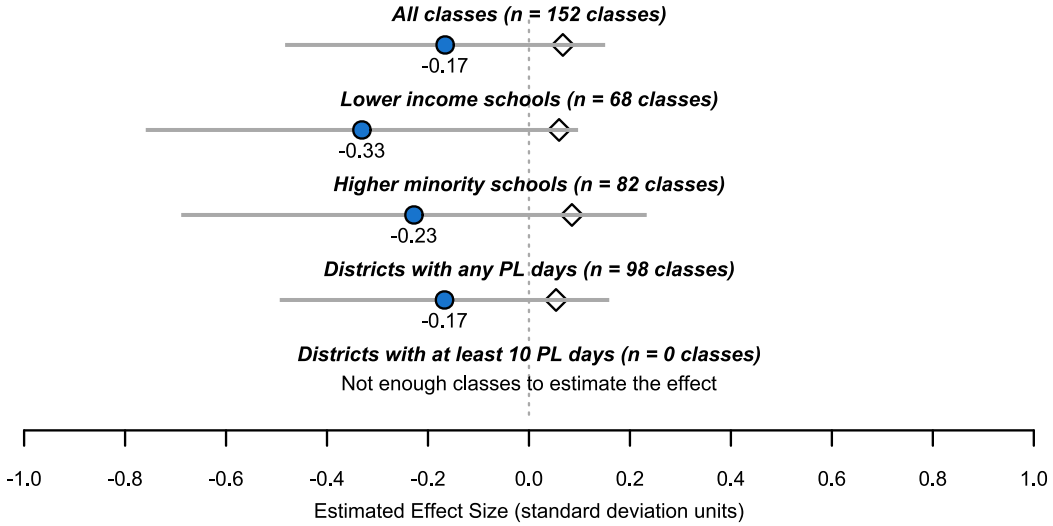
B. Effects over 3 years of access in grade-level classes with 3 years of access (*n* = 84 grade-level classes)



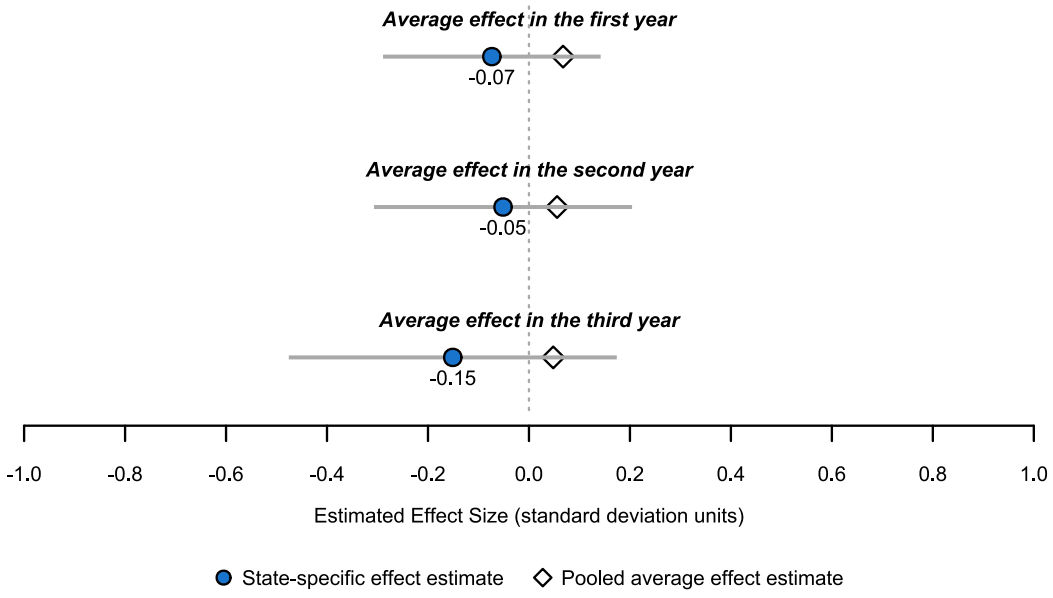
Notes. The exhibit presents the estimated effect (circles) of Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect. The diamonds depict the average effect across all states in the study.

Exhibit B.6. Estimated Effects in Wisconsin

A. Effects in the first year of access



B. Effects over 3 years of access in grade-level classes with 3 years of access (*n* = 302 grade-level classes)



Notes. The exhibit presents the estimated effect (circles) of Stepping Stones, where the effect is measured in standard deviations of the class-level percent proficient (log-odds) on the state's mathematics assessment. The horizontal gray lines represent the 95% confidence interval for the estimated effect. The diamonds depict the average effect across all states in the study.

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, with headquarters in Washington, D.C., the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance, both domestically and internationally, in the areas of education, health, and the workforce. For more information, visit www.air.org.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC | Austin, TX Arlington and Reston, VA | Seattle, WA

International: Algeria | Ethiopia | Germany | Haiti | Zambia

10732_02/20

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

The Effect of Stepping Stones Math on Student Achievement: Technical Supplement

FEBRUARY 2020

Jordan Rickles | So Jung Park

MAKING RESEARCH **RELEVANT**

The Effect of Stepping Stones Math on Student Achievement: Technical Supplement

FEBRUARY 2020

Jordan Rickles | So Jung Park

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Contents

Exhibits

Introduction

This document is a supplement to the final report for the Stepping Stones Math Impact Study: *The Effect of Stepping Stones Math on Student Achievement*. The technical supplement provides additional information about the methods used to estimate the effects of the Stepping Stones Math program.

Calculation of the Percent Proficient Outcome Metric

For the outcome measure, the American Institutes for Research (AIR) evaluation team calculated a standardized version of the mathematics percent proficient score for each class. The standardization included two steps. First, we transformed the percent proficient into a log-odds ratio (or logit) so that the measure would have properties of a normally distributed variable that are assumed for the standardization and linear models used for effect estimation:

$$logit_{ijgs} = log\left(\frac{pp_{ijgs}}{1-pp_{ijgs}}\right),$$

where pp_{ijgs} is the proportion of students for year *i* in class *j* within grade *g* in state *s* who scored at or above the state's proficiency benchmark on the state's mathematics assessment.

Second, we converted each class's logit score to a *z*-score metric based on the mean and standard deviation of logit scores within a year, grade, and state:

$$zscore_{ijgs} = \frac{logit_{ijgs} - \overline{logit_{igs}}}{SD(logit)_{igs}},$$

where $\overline{logit_{igs}}$ is the mean logit score across all classes within a particular year, grade, and state; and $SD(logit)_{igs}$ is the standard deviation of the logit scores across all classes within a particular year, grade, and state.

Detailed Matching Results

Matching within the comparative interrupted time series (CITS) design can help mitigate potential bias by comparing the outcome trajectories for treatment classes with a matched sample of comparison classes that experience similar external policy and contextual factors (e.g., a state changing the student assessment system or a regional economic downturn). Recent methodological studies demonstrate that CITS designs can produce treatment effect estimates equivalent to randomized controlled trials and regression discontinuity designs (e.g., Hallberg, Williams, & Swanlund, 2015; Jacob, Somers, Zhu, & Bloom, 2016).

To identify comparable classes, the AIR team matched each Stepping Stones class with a class in the same grade, state, location (i.e., urban, suburban, or town/rural), and school types (e.g., charter, magnet) that had no exposure to Stepping Stones. Among classes that met the criteria for these "exact" matches, we identified the most similar comparison class using an estimated propensity score. Classes with similar propensity scores should have, on average, similar prior achievement and characteristics. To test whether the matching process improved comparability of the Stepping Stones Math group and the matched comparison group, we compared the observed characteristics of the two groups.

A general rule of thumb for determining whether the two groups are similar is if the standardized mean difference (SMD) on measurable characteristics is less than 0.25 standard deviations (What Works Clearinghouse™, 2020).[1] The matching results pooled across the six-state sample are presented in the main report. In this supplement, we provide the matching results for each research question (See Exhibits 1 through 6) as well as the matching results for each state (See Exhibits 7 through 12).

[1] To calculate the SMD, we used Hedge's *g* for continuous measures and the Cox index for dichotomous measures.

Exhibit 1. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes: RQ 1 (All Classes)

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 454)	C Mean (*N* = 454)	SMD	C Mean (N = 117,910)	SMD
Location: City	0.44	0.44	0.00	0.36	0.22
Location: Suburb	0.39	0.39	0.00	0.39	0.00
Location: Town/Rural	0.17	0.17	0.00	0.25	-0.32
Magnet School	0.02	0.02	0.00	0.03	-0.25
Charter School	0.01	0.01	0.00	0.04	-0.74
Title I: Targeted	0.15	0.15	0.01	0.16	-0.05
Title I: Schoolwide	0.66	0.63	0.08	0.64	0.05
Total Enrollment (Grades 3–5)	494.48	524.57	-0.17	514.57	-0.09
Percent NSLP	0.51	0.49	0.09	0.59	-0.25
Percent Black	0.07	0.05	0.13	0.10	-0.18
Percent Asian/Pacific Islander	0.08	0.07	0.11	0.07	0.14
Percent Hispanic	0.30	0.30	-0.01	0.35	-0.14
Percent White	0.48	0.51	-0.10	0.43	0.15
Math Achievement: Prior Year 1 (*z*-score)	0.03	0.12	-0.09	0.00	0.03
Math Achievement: Prior Year 2 (*z*-score)	0.01	0.12	-0.10	0.00	0.01
Math Achievement: Prior Year 3 (*z*-score)	-0.06	0.02	-0.09	0.00	-0.06

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 2. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes: RQ 2 (Classes With 3 Years of Access)

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 447)	C Mean (*N* =447)	SMD	C Mean (N =117,910)	SMD
Location: City	0.44	0.44	0.00	0.36	0.20
Location: Suburb	0.36	0.36	0.00	0.39	-0.07
Location: Town/Rural	0.20	0.20	0.00	0.25	-0.19
Magnet School	0.02	0.02	0.00	0.03	-0.24
Charter School	0.01	0.01	0.00	0.04	-0.73
Title I: Targeted	0.19	0.18	0.04	0.16	0.13
Title I: Schoolwide	0.55	0.53	0.05	0.64	-0.21
Total Enrollment (Grades 3–5)	489.36	521.62	-0.18	514.57	-0.11
Percent NSLP	0.47	0.45	0.06	0.59	-0.42
Percent Black	0.06	0.05	0.13	0.10	-0.20
Percent Asian/Pacific Islander	0.08	0.07	0.11	0.07	0.15
Percent Hispanic	0.28	0.29	-0.03	0.35	-0.2
Percent White	0.51	0.54	-0.08	0.43	0.24
Math Achievement: Prior Year 1 (*z*-score)	0.16	0.22	-0.06	0.00	0.16
Math Achievement: Prior Year 2 (*z*-score)	0.12	0.21	-0.08	0.00	0.12
Math Achievement: Prior Year 3 (*z*-score)	0.07	0.09	-0.03	0.00	0.07

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 3. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes: RQ 3a (Classes in Lower Income Schools)

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 156)	C Mean (*N* = 156)	SMD	C Mean (N = 117,910)	SMD
Location: City	0.47	0.47	0.00	0.36	0.28
Location: Suburb	0.28	0.28	0.00	0.39	-0.32
Location: Town/Rural	0.26	0.26	0.00	0.25	0.02
Magnet School	0.01	0.01	0.00	0.03	-0.95
Charter School	0.01	0.01	0.00	0.04	-1.18
Title I: Targeted	0.01	0.03	-0.99	0.16	-2.06
Title I: Schoolwide	0.97	0.96	0.25	0.64	1.86
Total Enrollment (Grades 3–5)	470.81	499.92	-0.19	514.57	-0.19
Percent NSLP	0.72	0.72	0.00	0.59	0.48
Percent Black	0.10	0.08	0.12	0.10	-0.03
Percent Asian/Pacific Islander	0.04	0.03	0.12	0.07	-0.26
Percent Hispanic	0.39	0.41	-0.08	0.35	0.13
Percent White	0.42	0.42	-0.01	0.43	-0.03
Math Achievement: Prior Year 1 (*z*-score)	-0.50	-0.46	-0.05	0.00	-0.53
Math Achievement: Prior Year 2 (*z*-score)	-0.50	-0.48	-0.03	0.00	-0.52
Math Achievement: Prior Year 3 (*z*-score)	-0.57	-0.55	-0.03	0.00	-0.58

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 4. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes: RQ 3b (Classes in Higher Minority Schools)

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 255)	C Mean (*N* = 255)	SMD	C Mean (N = 117,910)	SMD
Location: City	0.65	0.65	0.00	0.36	0.72
Location: Suburb	0.33	0.33	0.00	0.39	-0.15
Location: Town/Rural	0.02	0.02	0.00	0.25	-1.71
Magnet School	0.03	0.03	0.00	0.03	0.03
Charter School	0.02	0.02	0.00	0.04	-0.39
Title I: Targeted	0.10	0.09	0.08	0.16	-0.32
Title I: Schoolwide	0.68	0.67	0.03	0.64	0.12
Total Enrollment (Grades 3–5)	501.04	548.76	-0.29	514.57	-0.06
Percent NSLP	0.55	0.56	-0.02	0.59	-0.12
Percent Black	0.07	0.07	0.05	0.10	-0.14
Percent Asian/Pacific Islander	0.11	0.10	0.07	0.07	0.36
Percent Hispanic	0.41	0.46	-0.19	0.35	0.20
Percent White	0.34	0.31	0.14	0.43	-0.29
Math Achievement: Prior Year 1 (*z*-score)	-0.11	-0.03	-0.09	0.00	-0.12
Math Achievement: Prior Year 2 (*z*-score)	-0.19	-0.10	-0.10	0.00	-0.20
Math Achievement: Prior Year 3 (*z*-score)	-0.23	-0.16	-0.07	0.00	-0.23

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 5. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes: RQ 4a (Classes in Districts Purchasing At Least 1 Day of Professional Learning)

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 356)	C Mean (*N* = 356)	SMD	C Mean (*N* = 117,910)	SMD
Location: City	0.54	0.54	0.00	0.36	0.46
Location: Suburb	0.38	0.38	0.00	0.39	-0.01
Location: Town/Rural	0.07	0.07	0.00	0.25	-0.88
Magnet School	0.03	0.03	0.00	0.03	-0.10
Charter School	0.01	0.01	0.00	0.04	-1.02
Title I: Targeted	0.12	0.15	-0.15	0.16	-0.21
Title I: Schoolwide	0.65	0.59	0.15	0.64	0.04
Total Enrollment (Grades 3–5)	505.78	532.54	-0.17	514.57	-0.04
Percent NSLP	0.53	0.50	0.11	0.59	-0.20
Percent Black	0.07	0.06	0.11	0.10	-0.16
Percent Asian/Pacific Islander	0.09	0.08	0.10	0.07	0.22
Percent Hispanic	0.34	0.34	0.00	0.35	-0.01
Percent White	0.43	0.46	-0.10	0.43	-0.01
Math Achievement: Prior Year 1 (*z*-score)	0.00	0.11	-0.12	0.00	0.00
Math Achievement: Prior Year 2 (*z*-score)	-0.03	0.09	-0.11	0.00	-0.03
Math Achievement: Prior Year 3 (*z*-score)	-0.12	-0.03	-0.10	0.00	-0.13

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 6. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes: RQ 4b (Classes in Districts Purchasing at Least 10 Days of Professional Learning)

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 207)	C Mean (*N* = 207)	SMD	C Mean (N = 117,910)	SMD
Location: City	0.66	0.66	0.00	0.36	0.75
Location: Suburb	0.30	0.30	0.00	0.39	-0.23
Location: Town/Rural	0.04	0.04	0.00	0.25	-1.29
Magnet School	0.04	0.04	0.00	0.03	0.24
Charter School	0.01	0.01	0.00	0.04	-0.68
Title I: Targeted	0.14	0.14	0.00	0.16	-0.13
Title I: Schoolwide	0.57	0.52	0.12	0.64	-0.18
Total Enrollment (Grades 3–5)	509.87	556.18	-0.29	514.57	-0.02
Percent NSLP	0.47	0.45	0.09	0.59	-0.40
Percent Black	0.04	0.04	-0.03	0.10	-0.32
Percent Asian/Pacific Islander	0.14	0.10	0.27	0.07	0.66
Percent Hispanic	0.39	0.38	0.04	0.35	0.14
Percent White	0.34	0.41	-0.27	0.43	-0.27
Math Achievement: Prior Year 1 (*z*-score)	0.17	0.30	-0.14	0.00	0.17
Math Achievement: Prior Year 2 (*z*-score)	0.16	0.22	-0.06	0.00	0.16
Math Achievement: Prior Year 3 (*z*-score)	0.13	0.14	-0.01	0.00	0.13

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 7. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes Pooled Across Cohorts: California

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 132)	C Mean (*N* = 132)	SMD	C Mean (*N* = 57,585)	SMD
Location: City	1.00	1.00	NA	0.42	NA
Location: Suburb	0.00	0.00	NA	0.43	NA
Location: Town/Rural	0.00	0.00	NA	0.14	NA
Magnet School	0.07	0.07	0.00	0.04	0.29
Charter School	0.05	0.05	0.00	0.08	-0.37
Title I: Targeted	0.09	0.04	0.56	0.11	-0.12
Title I: Schoolwide	0.55	0.55	-0.02	0.65	-0.26
Total Enrollment (Grades 3–5)	536.18	593.11	-0.32	582.05	-0.18
Percent NSLP	0.47	0.46	0.02	0.62	-0.54
Percent Black	0.03	0.02	0.07	0.05	-0.30
Percent Asian/Pacific Islander	0.20	0.13	0.41	0.10	0.72
Percent Hispanic	0.50	0.50	0.01	0.55	-0.19
Percent White	0.21	0.29	-0.41	0.25	-0.15
Math Achievement: Prior Year 1 (*z*-score)	0.20	0.34	-0.14	0.02	0.19
Math Achievement: Prior Year 2 (*z*-score)	0.08	0.15	-0.06	0.01	0.07
Math Achievement: Prior Year 3 (*z*-score)	0.06	0.10	-0.05	0.01	0.05

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 8. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes Pooled Across Cohorts: Illinois

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 76)	C Mean (*N* = 76)	SMD	C Mean (*N* = 21,532)	SMD
Location: City	0.00	0.00	NA	0.34	NA
Location: Suburb	0.96	0.96	0.00	0.44	2.08
Location: Town/Rural	0.04	0.04	0.00	0.22	-1.16
Magnet School	0.00	0.00	NA	0.04	NA
Charter School	0.00	0.00	NA	0.00	NA
Title I: Targeted	0.30	0.36	-0.14	0.31	-0.01
Title I: Schoolwide	0.57	0.51	0.13	0.50	0.16
Total Enrollment (Grades 3–5)	434.66	484.84	-0.30	465.05	-0.14
Percent NSLP	0.48	0.44	0.17	0.58	-0.31
Percent Black	0.04	0.04	-0.05	0.21	-0.53
Percent Asian/Pacific Islander	0.05	0.04	0.05	0.05	0.00
Percent Hispanic	0.34	0.38	-0.13	0.23	0.38
Percent White	0.55	0.50	0.18	0.48	0.21
Math Achievement: Prior Year 1 (*z*-score)	0.33	0.34	-0.01	0.02	0.32
Math Achievement: Prior Year 2 (*z*-score)	0.25	0.23	0.03	0.02	0.23
Math Achievement: Prior Year 3 (*z*-score)	-0.02	0.02	-0.06	0.01	-0.04

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to ONE because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 9. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes Pooled Across Cohorts: Kentucky

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 53)	C Mean (*N* = 53)	SMD	C Mean (*N* = 7,823)	SMD
Location: City	0.06	0.06	0.00	0.15	-0.64
Location: Suburb	0.06	0.06	0.00	0.19	-0.83
Location: Town/Rural	0.89	0.89	0.00	0.66	0.84
Magnet School	0.00	0.00	NA	0.02	NA
Charter School	0.00	0.00	NA	0.00	NA
Title I: Targeted	0.00	0.00	NA	0.03	NA
Title I: Schoolwide	0.94	0.96	-0.26	0.91	0.27
Total enrollment (Grades 3–5)	524.13	533.85	-0.05	471.60	0.32
Percent NSLP	0.57	0.55	0.17	0.62	-0.32
Percent Black	0.07	0.03	0.76	0.09	-0.14
Percent Asian/Pacific Islander	0.03	0.01	0.55	0.01	0.86
Percent Hispanic	0.11	0.06	0.43	0.06	0.66
Percent White	0.73	0.87	-0.86	0.80	-0.30
Math achievement: Prior Year 1 (*z*-score)	0.17	0.22	-0.07	0.00	0.17
Math achievement: Prior Year 2 (*z*-score)	0.12	0.28	-0.21	0.00	0.12
Math achievement: Prior Year 3 (*z*-score)	0.14	0.20	-0.07	0.00	0.14

Note. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 10. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes Pooled Across Cohorts: Ohio

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 40)	C Mean (*N* = 40)	SMD	C Mean (*N* = 14,210)	SMD
Location: City	0.33	0.33	0.00	0.26	0.20
Location: Suburb	0.38	0.38	0.00	0.37	0.02
Location: Town/Rural	0.30	0.30	0.00	0.37	-0.20
Magnet School	0.00	0.00	NA	0.00	NA
Charter School	0.00	0.00	NA	0.00	NA
Title I: Targeted	0.33	0.30	0.06	0.16	0.56
Title I: Schoolwide	0.40	0.43	-0.06	0.75	-0.91
Total Enrollment (Grades 3–5)	472.20	467.23	0.02	454.24	0.10
Percent NSLP	0.35	0.40	-0.21	0.57	-0.70
Percent Black	0.07	0.09	-0.2	0.20	-0.45
Percent Asian/Pacific Islander	0.05	0.03	0.44	0.02	0.96
Percent Hispanic	0.08	0.09	-0.14	0.06	0.18
Percent White	0.74	0.75	-0.05	0.67	0.22
Math Achievement: Prior Year 1 (*z*-score)	0.47	0.42	0.09	-0.10	0.57
Math Achievement: Prior Year 2 (*z*-score)	0.47	0.35	0.21	-0.10	0.57
Math Achievement: Prior Year 3 (*z*-score)	0.60	0.61	-0.01	-0.08	0.69

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 11. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes Pooled Across Cohorts: Washington

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 77)	C Mean (*N* = 77)	SMD	C Mean (*N* = 7,295)	SMD
Location: City	0.26	0.26	0.00	0.41	-0.42
Location: Suburb	0.61	0.61	0.00	0.36	0.63
Location: Town/Rural	0.13	0.13	0.00	0.23	-0.42
Magnet School	0.00	0.00	NA	0.00	NA
Charter School	0.00	0.00	NA	0.00	NA
Title I: Targeted	0.12	0.17	-0.24	0.12	0.00
Title I: Schoolwide	0.81	0.70	0.31	0.65	0.47
Total Enrollment (Grades 3–5)	549.05	548.12	0.01	469.95	0.57
Percent NSLP	0.59	0.53	0.28	0.50	0.39
Percent Black	0.03	0.03	0.00	0.05	-0.24
Percent Asian/Pacific Islander	0.04	0.06	-0.37	0.08	-0.38
Percent Hispanic	0.26	0.23	0.20	0.23	0.18
Percent White	0.54	0.57	-0.13	0.54	0.04
Math Achievement: Prior Year 1 (*z*-score)	-0.27	-0.13	-0.17	0.03	-0.30
Math Achievement: Prior Year 2 (*z*-score)	-0.18	-0.03	-0.15	0.03	-0.21
Math Achievement: Prior Year 3 (*z*-score)	-0.33	-0.28	-0.05	0.02	-0.35

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Exhibit 12. Characteristics of Matched Stepping Stones Math Classes and Comparison Classes Pooled Across Cohorts: Wisconsin

Characteristic	Matched Classes			All Eligible Comparison Classes	
	SS Mean (*N* = 166)	C Mean (*N* = 166)	SMD	C Mean (*N* = 9,465)	SMD
Location: City	0.39	0.39	0.00	0.27	0.32
Location: Suburb	0.46	0.46	0.00	0.24	0.61
Location: Town/Rural	0.16	0.16	0.00	0.50	-1.01
Magnet School	0.00	0.00	NA	0.00	NA
Charter School	0.02	0.02	0.00	0.05	-0.67
Title I: Targeted	0.27	0.25	0.04	0.30	-0.12
Title I: Schoolwide	0.44	0.42	0.04	0.47	-0.07
Total Enrollment (Grades 3–5)	422.60	456.64	-0.22	377.18	0.29
Percent NSLP	0.42	0.42	0.00	0.45	-0.11
Percent Black	0.11	0.09	0.16	0.07	0.21
Percent Asian/Pacific Islander	0.02	0.05	-0.59	0.04	-0.31
Percent Hispanic	0.16	0.19	-0.19	0.11	0.29
Percent White	0.67	0.63	0.16	0.73	-0.24
Math Achievement: Prior Year 1 (*z*-score)	-0.01	0.03	-0.04	0.06	-0.08
Math Achievement: Prior Year 2 (*z*-score)	-0.01	0.13	-0.13	0.07	-0.08
Math Achievement: Prior Year 3 (*z*-score)	0.03	0.11	-0.08	0.06	-0.03

Notes. The statistics for all eligible comparison classes provide a general reference for how the Stepping Stones Math classes compare to the broader population of eligible classes and how matching improved comparability. The four race/ethnicity composition measures do not sum to one because the Native American and "multiple race" categories are not reported in the table.

SS = Stepping Stones Math class; C = comparison class; SMD = standardized mean difference; NSLP = National School Lunch Program.

Impact Model Details

RQ 1: The Effect of Stepping Stones Math Access in the First Year

For the first research question (RQ 1), we estimated a baseline trend model with an effect estimate for the first year of program access (i.e., the cohort start year). To estimate the baseline trend model, we used the following hierarchical linear model, estimated separately for each state:

(3.1) Level 1 (Year):

$$Y_{ij} = \beta_{0j} + \beta_{1j} RELYEAR_{ij} + \beta_{2j} FY1_{ij} + e_{ij}$$

(3.2) Level 2 (Class):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} SS_j + \gamma_{02} GRADE_j + \gamma_{03} COHORT_j + \gamma_{04} Z_j + u_{0k}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} SS_j$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} SS_j$$

(3.3) Combined Model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} SS_j + \gamma_{10} RELYEAR_{ij} + \gamma_{11} (RELYEAR_{ij} \times SS_j) + \gamma_{20} FY1_{ij} + \gamma_{21} (FY1_{ij} \times SS_j) + \gamma_{02} GRADE_j + \gamma_{03} COHORT_j + \gamma_{04} Z_j + e_{ij} + u_{0j}$$

where *i* indexes years and *j* indexes classes.

In this model, Y_{ij} is the mathematics percent proficient (standardized) in year *i* for class *j*.[2] $RELYEAR_{ij}$ is a continuous variable for time period centered at the last baseline year. $FY1_{ij}$ is a dichotomous indicator for whether the outcome measure was in the first follow-up year (i.e., the cohort start year). SS_j is a dichotomous indicator for whether the class had access to Stepping Stones Math ($SS_j = 1$) or whether the class was a comparison class ($SS_j = 0$). The model includes controls for a vector of dichotomous indicators for each class's grade level ($GRADE_j$), a vector of dichotomous indicators for each class's cohort ($COHORT_j$), and a vector of time-constant school characteristics (Z_j). The measures included in the time-constant school characteristic vectors are the measures included in the matching approach (see Appendix A, Exhibit A.1, in the final report).

[2] To minimize the influence of outliers, we limited Y_{ij} (the standardized mathematics percent proficient) within the range between -3 and 3. All data below -3 were recoded to -3, and data above 3 were recoded to 3.

With this model specification, the primary parameter of interest is γ_{21}, which represents the post-treatment differences in outcome trend between comparison and Stepping Stones classes for the first post-treatment year after adjusting for any differences in school characteristics.

To estimate the model, we used estimation procedures that adjust the standard errors for the autoregressive nature of the annual outcome measures and the additional clustering of classes within schools.

To summarize model estimates across each state, we calculated precision-weighted averages, where precision is defined as the inverse of the variance.

RQ 2: The Effect of Stepping Stones Math Access Across 3 Years

For RQ 2, we restricted the analysis to classes that had access to Stepping Stones Math for 3 years (and their matched comparison classes) and estimated a baseline trend model that parallels the structure of the model used for RQ 1. To estimate the average effect in each of the 3 years, we used the following hierarchical linear model, estimated separately for each state:

(4.1) Level 1 (Year):

$$Y_{ij} = \beta_{0j} + \beta_{1j}RELYEAR_{ij} + \beta_{2j}FY1_{ij} + \beta_{3j}FY2_{ij} + \beta_{4j}FY3_{ij} + e_{ij}$$

(4.2) Level 2 (Class):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}SS_j + \gamma_{02}GRADE_j + \gamma_{03}COHORT_j + \gamma_{04}Z_j + u_{0k}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}SS_j$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}SS_j$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}SS_j$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}SS_j$$

(4.3) Combined Model:

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{01}SS_j + \gamma_{10}RELYEAR_{ij} + \gamma_{20}FY1_{ij} + \gamma_{30}FY2_{ij} + \gamma_{40}FY3_{ij} \\ & + \gamma_{11}\left(RELYEAR_{ij} \times SS_j\right) + \gamma_{21}\left(FY1_{ij} \times SS_j\right) + \gamma_{31}\left(FY2_{ij} \times SS_j\right) + \gamma_{41}\left(FY3_{ij} \times SS_j\right) \\ & + \gamma_{02}GRADE_j + \gamma_{03}COHORT_j + \gamma_{04}Z_j + e_{ij} + u_{0j} \end{aligned}$$

where *i* indexes years and *j* indexes classes.

The terms in this model are defined the same as in Equation 3. However, in addition to the dichotomous indicator for whether the outcome was in the first follow-up year ($FY1_{ij}$), this model includes indicators for whether the outcome was in the second and third follow-up years

($FY2_{ij}$, $FY3_{ij}$). This enables us to estimate the average effect of Stepping Stones Math during the first post-treatment year (γ_{21}), the second post-treatment year (γ_{31}), and the third post-treatment year (γ_{41}), respectively.

RQ 3 and RQ 4: The Effect of Stepping Stones Math in Different School and District Settings in the First Year

For the RQ 3 and RQ 4, we estimated a baseline trend model identical to that for Equation 3, but restricted the analysis to different subsamples. Specifically, we estimated two separate models for each research question. For RQ 3, one analysis was restricted to classes in schools with more than half of their students eligible for free or reduced-price lunch (and their matched comparison classes), and the other analysis was restricted to classes in schools with at least a third of their students classified as a non-White racial/ethnic minority group (and their matched comparison classes). For RQ 4, one analysis was restricted to classes in districts with at least 1 day of professional learning (and their matched comparison classes), and the other analysis was restricted to classes in schools with at least 10 days of professional learning (and their matched comparison classes).

References

Hallberg, K., Williams, R. T., & Swanlund, A. (November 2015). *Assessing the validity of comparative interrupted time series designs in practice: Lessons learned from two within-study comparisons*. Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Miami, FL.

Jacob, R., Somers, M. A., Zhu, P., & Bloom, H. (2016). The validity of the comparative interrupted time series design for evaluating the effect of school-level interventions. *Evaluation Review, 40*(3), 167–198.

What Works Clearinghouse. (2020). S*tandards handbook* (*version 4.1*). Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, with headquarters in Washington, D.C., the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance, both domestically and internationally, in the areas of education, health, and the workforce. For more information, visit www.air.org.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC | Austin, TX Arlington and Reston, VA | Seattle, WA

International: Algeria | Ethiopia | Germany | Haiti | Zambia